

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:54:19 ; Search time 2613.74 Seconds
(without alignments)
192.153 Million cell updates/sec

Title: US-09-986-381-5
Perfect score: 24
Sequence: 1 tgcctgacttcaactctgtctc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

C	1	24	100.0	159	6	A39493	A39493 Sequence 3
	2	24	100.0	600	6	I92482	I92482 Sequence 2
	3	24	100.0	799	9	HSP5314	X92659 H.sapiens 1
	4	24	100.0	3407	9	HOMOTSP1	AF136270 Homo sapi
	5	24	100.0	3423	9	HSM059JP1	AF135120 Homo sapi
	6	24	100.0	4360	9	AB018045	AB018045 Homo sapi
	7	24	100.0	20303	9	HSP53G	X54156 Human p53 g
	8	24	100.0	20303	9	HSU94788	U94788 Human p53 (
	9	24	100.0	159021	2	AC087388	AC087388 Homo sapi
	10	24	100.0	160457	2	AC008049	AC008049 Homo sapi
	11	22	91.7	121	6	AX262802	AX262802 Sequence
C	12	22	91.7	121	6	AX262803	AX262803 Sequence
	13	19.4	80.8	2457	5	AY062922	AY062922 Xenopus 1
	14	19.4	80.8	11086	9	AL137075	AL137075 Human DNA
C	15	19.2	80.0	45545	2	AC095545	AC095545 Rattus no
	16	19.2	80.0	66569	2	AC103837	AC103837 Homo sapi
C	17	19.2	80.0	109343	9	AC007161	AC007161 Homo sapi
	18	19.2	80.0	125133	9	AC013274	AC013274 Homo sapi
	19	19.2	80.0	137074	9	AF130248	AF130248 Homo sapi
C	20	19.2	80.0	157875	9	HS272L16	AL023754 Human DNA
	21	19.2	80.0	162813	2	AC021833	AC021833 Homo sapi
C	22	19.2	80.0	167799	2	AC090881	AC090881 Mus muscu
	23	19.2	80.0	171915	9	AF246928	AF246928 Homo sapi
	24	19.2	80.0	187928	2	AC093624	AC093624 Homo sapi
C	25	19.2	80.0	190006	2	AC090499	AC090499 Homo sapi
	26	19.2	80.0	190686	2	AC093731	AC093731 Homo sapi
	27	19.2	80.0	197126	2	AC023130	AC023130 Homo sapi
	28	19.2	80.0	200920	2	AC069319	AC069319 Homo sapi
	29	19.2	80.0	201380	2	AC092905	AC092905 Homo sapi
	30	19.2	80.0	208325	2	AC091682	AC091682 Mus muscu
C	31	19.2	80.0	222895	2	AC084065	AC084065 Mus muscu
	32	19.2	80.0	340000	9	HS21C008	AL163208 Homo sapi
	33	19	79.2	23	6	AX137493	AX137493 Sequence
	34	19	79.2	23	6	BD006891	BD006891 Continuo
	35	18.8	78.3	716	9	HS327499	AF327499 Homo sapi
C	36	18.8	78.3	11000	3	AF297644	AF297644 Drosophil
C	37	18.8	78.3	33892	2	AC019487	AC019487 Drosophil
	38	18.8	78.3	112604	2	AC095989	AC095989 Rattus no
	39	18.8	78.3	133292	2	OSJN00152	AL662950 Oryza sat
	40	18.8	78.3	165700	2	AC099450	AC099450 Rattus no
C	41	18.8	78.3	169048	9	AL158011	AL158011 Human DNA
C	42	18.8	78.3	186686	3	AC008192	AC008192 Drosophil
C	43	18.8	78.3	188608	9	CNS00005	AL049779 Human chr
	44	18.8	78.3	191896	9	AC009481	AC009481 Homo sapi
	45	18.8	78.3	224774	2	AL645646	AL645646 Mus muscu

ALIGNMENTS

RESULT	1	A39493	Sequence 3 from Patent WO9416066.	159 bp	DNA	linear	PAT 05-MAR-1997
A39493/c		A39493					
LOCUS		A39493					
DEFINITION		A39493					
ACCESSION		A39493.1	GI:2295812				
VERSION							
KEYWORDS							
SOURCE		human.					
ORGANISM		Homo sapiens					
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE		1 (bases 1 to 159)					
AUTHORS		Bergmann, J.E. and Preddie, R.E.					
TITLE		AGENTS FOR THE PREVENTION AND TREATMENT OF HIV REPLICATION AND AIDS IN HUMANS					
JOURNAL		Patent: WO 9416066-A 3 21-JUL-1994;					
COMMENT		BERGMANN JOHANNA E (DE)					
		Other publication AU 5969094 940815					
		Other publication CA 2086577 940701.					
FEATURES		Location/Qualifiers					
source		1..159					
		/organism="Homo sapiens"					
		/db_xref="taxon:9606"					

BASE COUNT ORIGIN	44 a	29 c	62 g	24 t	
Query Match	100.0%; Score 24; DB 6; Length 159;				
Best Local Similarity	100.0%; Pred No. 0.42; Length 159;				
Matches	24; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
Qy	1 tgcctgacttcaactctgtctc 24				
Db	144 TGCCTGACTTTCAACTCTGTCTC 121				
RESULT 2					
LOCUS	I92482 600 bp DNA linear PAT 01-DEC-1998				
DEFINITION	Sequence 2 from patent US 5728526.				
ACCESSION	I92482				
VERSION	I92482.1 GI:3936952				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 600)				
AUTHORS	George, A.L. Jr., Bhatnagar, S.K. and Nazarenko, I.				
TITLE	Method for analyzing a nucleotide sequence				
JOURNAL	Patent: US 5728526-A 2 17-MAR-1998;				
FEATURES	Location/Qualifiers				
source	1..600				
BASE COUNT	102 a	170 c	168 g	160 t	
ORIGIN					
Query Match	100.0%; Score 24; DB 6; Length 600;				
Best Local Similarity	100.0%; Pred. No. 0.38;				
Matches	24; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
Qy	1 tgcctgacttcaactctgtctc 24				
Db	65 TGCCTGACTTTCAACTCTGTCTC 88				
RESULT 3					
HSP5314					
LOCUS	HSP5314 799 bp DNA linear PRI 25-NOV-1996				
DEFINITION	H.sapiens intron 4 from p53 gene.				
ACCESSION	X92659				
VERSION	X92659.1 GI:1177472				
KEYWORDS	p53 gene.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Shamsher, M. and Montano, X.				
TITLE	Analysis of intron 4 of the p53 gene in human cutaneous melanoma				
JOURNAL	Gene 176 (1-2), 259-262 (1996)				
MEDLINE	97075940				
REFERENCE	2 (bases 1 to 799)				
AUTHORS	Montano, X.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-OCT-1995) X.C. Montano, Imperial Cancer Research Fund, Viral Mediated Cell Differentiation Lab, PO Box 123, Lincoln's Inn Fields, LONDON, WC2A 3PX, UK				
FEATURES	Location/Qualifiers				
source	1..799				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="17p"				
	/cell_line="ht6.1, mjm, sk23, dx3, hmb-2, a375p, a375m, ht1080, mrc5"				
	/cell_type="melanoma"				

exon	variation	/gene="p53" /number=8 2987 /gene="p53" /replace="g" 3199..3272 /gene="p53" /number=9		
BASE COUNT	738 a	963 c	877 g	829 t
ORIGIN				
Query Match 100.0%; Score 24; DB 9; Length 3407;				
Best Local Similarity 100.0%; Pred. No. 0.34;				
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	tgccctgactttcaactctgtctc 24		
Db	1515	TGCCCTGACTTTCAACTCTGTCTC 1538		
RESULT	5			
LOCUS	HSM059JP1	3423 bp	DNA	linear
DEFINITION	Homo sapiens tumor suppressor protein p53 (p53) gene, exons 2 through 9.			
ACCESSION	AF135120			
VERSION	AF135120.1	GI:4731629		
KEYWORDS	1 of 2			
SEGMENT	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	Allalunis-Turner,M.J., Barron,G.M., Day,R.S. III, Dobler,K.D. and Mirzayans,R.			
AUTHORS	1 (bases 1 to 3423)			
TITLE	Isolation of two cell lines from a human malignant glioma specimen differing in sensitivity to radiation and chemotherapeutic drugs			
JOURNAL	Radiat. Res. 134 (3), 349-354 (1993)			
MEDLINE	93303270			
REFERENCE	2 (bases 1 to 3423)			
AUTHORS	Anderson,C.W., Kieletzawa,J. and Allalunis-Turner,M.J.			
TITLE	Human p53 from malignant glioma-derived M059J and M059K cells has a cancer-associated mutation in exon 8			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 3423)			
AUTHORS	Anderson,C.W., Kieletzawa,J. and Allalunis-Turner,M.J.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-MAR-1999) Biology, Brookhaven National Laboratory, 50 Bell Avenue, Upton, NY 11973-5000, USA			
FEATURES	Location/Qualifiers			
Source	1..3423			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/cell_type="M059J"			
	/note="M059J cells produce a mutant form of p53"			
exon	<216..289			
	/gene="p53"			
	/number=2			
exon	407..428			
	/gene="p53"			
	/number=3			
exon	538..816			
	/gene="p53"			
	/number=4			
exon	1571..1754			
	/gene="p53"			
	/number=5			
exon	1836..1948			
	/gene="p53"			
	/number=6			
exon	2517..2626			
	/gene="p53"			
	/number=7			
exon	2970..3106			

```

exon          3443..3954
/gene="HSP70-1"
/note="alternative splicing
transcription usually starts from exon 2"
number=2
3955..>4360
/gene="HSP70-1"
/note="alternative splicing"
/number=3
BASE COUNT    970 a 1242 c 1147 g 1001 t
ORIGIN

Query Match      100.0%; Score 24; DB 9; Length 4360;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcctgactttcaactctgtctc 24
|||||
Db 717 TGCCTGACTTTCACACTCTGTCTC 740

RESULT 7
HSP53G      20303 bp      DNA      linear      PRI 25-JUN-1997
LOCUS       Human p53 gene for transformation related protein p53 (also called
DEFINITION  transformation-associated protein p53, cellular tumor antigen p53,
and non-viral tumour antigen p53).
ACCESSION   X54156
VERSION     X54156.1 GI:35213
KEYWORDS    anti-oncogene; cell cycle control; growth suppressor; heat shock
protein 70; oncogene; p53 cellular tumour antigen; p53 gene;
phosphoprotein; transforming capacity; tumor antigen.
SOURCE      human
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 20303)
AUTHORS     Chumakov,P.M.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-1990) Chumakov P.M., Engelhardt Inst. of
Molecular Biology, Academy of Science of the USSR, Vavilov St. 32,
117984 Moscow, USSR
REFERENCE   2 (bases 1 to 20303)
AUTHORS     Chumakov,P.M., Almazov,V.P. and Jenkins,J.R.
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 20303)
AUTHORS     Futreal,P.A., Barrett,J.C. and Wiseman,R.W.
TITLE       An Alu polymorphism intragenic to the TP53 gene
JOURNAL     Nucleic Acids Res. 19 (24), 6977 (1991)
MEDLINE     92107726
COMMENT     See also entries K03199, M14690, M14695, X01405, X02469, M22881-4,
M22887-8, M22894-8.
See also Mol. Cell. Biol. 6:1379-1385(1986);
and Mol. Cell. Biol. 7:961-963(1987).

FEATURES
source      1..20303
            /organism="Homo sapiens"
            /strain="caucasian"
            /db_xref="taxon:9606"
            /chromosome="17"
            /map="p13"
mRNA        join(843..949,11689..11790,11906..11927,12021..12299,
13055..13238,13320..13432,14000..14109,14452..14588,
14681..14754,17572..17678,18599..19876)
            /gene="p53"
prim_transcript 843..19876
            /gene="p53"
exon        843..949
            /gene="p53"
            /number=1
gene         843..19876
            /gene="p53"

950..11688
/gene="p53"
/number=1
2581..2587
/gene="p53"
/note="5'-ALU flanking"
2588..2877
/gene="p53"
/rpt_family="ALU"
2890..2896
/gene="p53"
/note="3' ALU-flanking"
3915..3929
/gene="p53"
/note="3' ALU-flanking"
3950..4223
/gene="p53"
/rpt_family="ALU"
4224..4238
/gene="p53"
/note="5'-ALU flanking"
4319..4327
/gene="p53"
/note="5'-ALU flanking"
4328..4603
/gene="p53"
/rpt_family="ALU"
4631..4639
/gene="p53"
/note="3' ALU-flanking"
4786..5574
/gene="p53"
/note="rearranged cluster"
5802..5811
/gene="p53"
/note="5'-ALU flanking"
5812..6100
/gene="p53"
/rpt_family="ALU"
6127..6136
/gene="p53"
/note="3' ALU-flanking"
6221..6236
/gene="p53"
/note="5'-ALU flanking"
6237..6517
/gene="p53"
/rpt_family="ALU"
6531..6546
/gene="p53"
/note="3' ALU-flanking"
6548..7812
/gene="p53"
/note="rearranged cluster"
/rpt_family="ALU"
8703..8982
/gene="p53"
/rpt_family="ALU"
9087..9098
/gene="p53"
/note="5'-ALU flanking"
9099..9377
/gene="p53"
/rpt_family="ALU"
9391..9402
/gene="p53"
/note="3' ALU-flanking"
9513..10332
/gene="p53"
/note="rearranged cluster"
/rpt_family="ALU"
11065..11069

```


Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., Laroque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 13, 2001 this sequence version replaced gi:16041379.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11969

Center clone name: 199_F_11

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 5823 5922: contig of 5822 bp in length
* 5823 5922: gap of 100 bp
* 5923 43817: contig of 37895 bp in length
* 43818 43917: gap of 100 bp
* 43918 119706: contig of 75789 bp in length
* 119707 119806: gap of 100 bp
* 119807 153208: contig of 33402 bp in length
* 153209 153308: gap of 100 bp
* 153309 159021: contig of 5713 bp in length.

FEATURES

source

1. .159021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17"
/clone="RP11-199F11"
/clone_lib="RP11-11 Human Male BAC"
BASE COUNT 40373 a 39760 c 38883 g 39366 t 639 others
ORIGIN

Query Match 100.0%; Score 24; DB 2; Length 159021;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcctgaacttcaactctgtctc 24
|||||

Db 78342 TGCCCTGACTTTCACACTCTCTC 78365

RESULT 10

AC008049

LOCUS

DEFINITION

AC008049 Homo sapiens clone RP11-199F11, WORKING DRAFT SEQUENCE, 9 unordered

pieces.

AC008049.37 GI:14787096

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.

human.

AC008049 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 160457)

Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,

Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtege,O., Lieu,C., Liu,J., Liu,W.,

Lozsed,H., Lozsed,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Ogub,M., Okwunodu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,

Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,

Ston,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,P., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 160457)

Worley,K.C.

Direct Submission

Submitted (16-JUL-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 17, 2001 this sequence version replaced gi:14328991.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HMGY

Center clone name: RP11-199F11

----- Summary Statistics

Sequencing vector: Plasmid; M77789

Sequencing vector: M13; L08821

Chemistry: Dye-primer Bodipy: 30% of reads

Chemistry: Dye-terminator Big Dye: 66% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 161924 bases at least Q40
 Consensus quality: 166972 bases at least Q30
 Consensus quality: 169776 bases at least Q20
 Estimated insert size: 162538; sum-of-contigs estimation
 Quality coverage: 10.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 55987: contig of 55987 bp in length
 * 55988 56087: gap of unknown length
 * 56088 96796: contig of 40709 bp in length
 * 96797 96896: gap of unknown length
 * 96897 124682: contig of 27786 bp in length
 * 124683 124782: gap of unknown length
 * 124783 136187: contig of 11404 bp in length
 * 136188 136286: gap of unknown length
 * 136287 141789: contig of 5503 bp in length
 * 141790 141889: gap of unknown length
 * 141890 149503: contig of 7614 bp in length
 * 149504 149603: gap of unknown length
 * 149604 154667: contig of 5064 bp in length
 * 154668 154767: gap of unknown length
 * 154768 156994: contig of 2227 bp in length
 * 156995 157094: gap of unknown length
 * 157095 160457: contig of 3363 bp in length.
 * Location/Qualifiers
 * source
 * 1. .160457

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-199F11"

BASE COUNT 39372 a 39284 c 40113 g 40841 t 847 others
 ORIGIN

Query Match 100.0%; Score 24; DB 2; Length 160457;
 Best Local Similarity 100.0%; Pred. No. 0.26;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcctgactttcaactctgtctc 24
 |||||
 Db 72676 TGCCCTGACTTTCAACTCTGTCTC 72699

RESULT 11
 AX262802
 LOCUS AX262802 121 bp DNA linear PAT 26-OCT-2001
 DEFINITION Sequence 193 from Patent W00173002.
 ACCESSION AX262802
 VERSION AX262802.1 GI:16511601
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)

REFERENCE
 AUTHORS Kniec,E.B., Gamper,H.B. and Rice,M.C.
 TITLE Targeted chromosomal genomic alterations with modified single
 stranded oligonucleotides

JOURNAL Patent: WO 0173002-A 193 04-OCT-2001;
 UNIVERSITY OF DELAWARE (US)

FEATURES
 Location/Qualifiers
 source
 1. .121
 /organism="Homo sapiens"

BASE COUNT 20 a 48 c 20 g 33 t
 ORIGIN

Query Match 91.7%; Score 22; DB 6; Length 121;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cccgtgactttcaactctgtctc 24
 |||||
 Db 1 CCCTGACTTTCAACTCTGTCTC 22

RESULT 12
 AX262803/c
 LOCUS AX262803 121 bp DNA linear PAT 26-OCT-2001
 DEFINITION Sequence 194 from Patent W00173002.
 ACCESSION AX262803
 VERSION AX262803.1 GI:16511602
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)

REFERENCE
 AUTHORS Kniec,E.B., Gamper,H.B. and Rice,M.C.
 TITLE Targeted chromosomal genomic alterations with modified single
 stranded oligonucleotides

JOURNAL Patent: WO 0173002-A 194 04-OCT-2001;
 UNIVERSITY OF DELAWARE (US)

FEATURES
 Location/Qualifiers
 source
 1. .121
 /organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 33 a 20 c 48 g 20 t
 ORIGIN

Query Match 91.7%; Score 22; DB 6; Length 121;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cccgtgactttcaactctgtctc 24
 |||||
 Db 121 CCCTGACTTTCAACTCTGTCTC 100

RESULT 13
 AX262922
 LOCUS AX262922 2457 bp mRNA linear VRT 03-FEB-2002
 DEFINITION Xenopus laevis phosphoinositide-3-kinase (PIK3R1) mRNA, complete
 cds.
 ACCESSION AY062922
 VERSION AY062922.1 GI:18482451
 KEYWORDS
 SOURCE African clawed frog.

ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 2457)

REFERENCE
 AUTHORS Al-Khalili,O.K. and Eaton,D.C.
 TITLE Molecular cloning of Xenopus laevis phosphoinositide-3-kinase
 (regulatory subunit, polypeptide 1, P85 alpha)

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2457)

AUTHORS Al-Khalili,O.K. and Eaton,D.C.
 TITLE Direct Submission

JOURNAL Submitted (15-NOV-2001) Physiology, Emory University, 1648 Pierce
 Dr, Atlanta, GA 30322, USA

FEATURES
 Location/Qualifiers
 source
 1. .2457

```
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_line="A6"
1..2457
/gene="PIK3R1"
242..2410
/gene="PIK3R1"
/notes="protein kinase; p85 alpha; regulatory subunit;
polyptide 1"
/codon_start=1
/product="phosphoinositide-3-kinase"
/protein_id="AAL68953.1"
/db_xref="GI:1848252"
/translation="MSADGKYKRALPYKVEREDIDLVRGDTLSVSRATLALGCPA
EGMEARPEQIGLWNCNENTGCGDFCTGYVYGLGRKISPTPKPRPLVPAPGA
ARGEAAQAFPLDLAEQFSPDIPGVPLIKLIEAKKGLAESATLRSQSSSLTEL
RQLKCDAAVDFQFDIALDALKRYFLDLPHPVIPALMYSEMLSAARESSSDRY
VOLLKIIIRCANVQOQYWLTLQYLKHFFQPCQSSKNLLSAKSLAETFPFLFRIOV
SSDSSESTQILEVLIASEWNEKQAPALPKPKSKSTAVTNNGNNKLIFHRDQKGFPS
GDISREEVNELRDTTGTFLVRDASTIMHGDTLTLRKGGNNKLIFHRDQKGFPS
DPLTFNSVELITHYRNESLAQYNPKLDVLLYVSRQDOOVVKESIDAVGKKLRE
YNVOFEENQEDRLYEDYTTTSQEIOMKRTAIAFNETIKFEEQCCQEKYSKEIK
EKFRREGNEKEQRLHNYEKLSKRSISEIVDSKRLEEDLKQTAENREIDKRNNSIK
PDLLQRLTRDOYLMWLTQKGVQKKNLWNGNENTEDQYSVVEEDLPQODERTWV
GNINRNQAEINLRKRDGTFLVRSSKAGCFACSVMAEGEVKHCVINKTHTGFGFAEP
YNLYSSLKELVLHYQYTSLVQHNDLSNVLTAHPVYAAQRR"
BASE COUNT 733 a 566 c 624 g 534 t
ORIGIN

Query Match 80.8%; Score 19.4; DB 5; Length 2457;
Best Local Similarity 95.2%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcctgacttcaactctgt 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1420 TCCCTGACTTTTCAACTCTGT 1440

RESULT 14
AL137075/c
LOCUS
DEFINITION
AL137075
ACCESSION
AL137075
VERSION
AL137075.22 GI:10334452
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Coville,G.
Direct Submission
Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 27, 2000 this sequence version replaced gi:10119670.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1

IMPORTANT: This sequence is not the entire insert of clone
RP5-1082D3 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP4-738P11 is at 10987 in this sequence.
The true right end of clone RP5-968D22 is at 100 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
RP5-1082D3 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
Location/Qualifiers
1..11086
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q23.1-24.3"
/clone="RP5-1082D3"
/clone_lib="RPCI-5"
9..164
repeat_region
/notes="MIR repeat: matches 2..156 of consensus"
repeat_region
1095..1191
/notes="L2 repeat: matches 2613..2707 of consensus"
repeat_region
1356..1493
/notes="MIR repeat: matches 3..150 of consensus"
repeat_region
1497..1630
/notes="MER91A repeat: matches 49..196 of consensus"
repeat_region
2812..3065
/notes="MIR repeat: matches 2..259 of consensus"
misc_feature
3850..4458
/notes="match: GSS: Em:AQ280853"
repeat_region
4925..4956
/notes="16 copies 2 mer ac 90% conserved"
misc_feature
6051..6531
/notes="match: GSS: Em:AQ246599"
repeat_region
6998..7272
/notes="L2 repeat: matches 2442..2750 of consensus"
repeat_region
7530..7638
/notes="MIR repeat: matches 88..191 of consensus"
repeat_region
7899..7983
/notes="MIR repeat: matches 57..145 of consensus"
repeat_region
8686..8808
/notes="MIR repeat: matches 14..145 of consensus"
repeat_region
9952..10031
/notes="L2 repeat: matches 2575..2664 of consensus"
repeat_region
10446..10871
/notes="L2 repeat: matches 825..1305 of consensus"
repeat_region
10932..10956
/notes="L2 repeat: matches 1321..1345 of consensus"
repeat_region
10958..11085
/notes="MER94 repeat: matches 13..134 of consensus"
BASE COUNT 2951 a 2368 c 2414 g 3353 t
ORIGIN

Query Match 80.8%; Score 19.4; DB 9; Length 11086;
Best Local Similarity 95.2%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcctgacttcaactctgt 21
| | | | | | | | | | | | | | | | | | | | | |
Db 8300 TGCCCTGCTTCAACTCTGT 8280

RESULT 15
AC095545
LOCUS
DEFINITION
Rattus norvegicus clone CH230-8M11, *** SEQUENCING IN PROGRESS ***
AC095545
ACCESSION
AC095545
```

```

VERSION AC095545.2 GI:17942067
SOURCE HTG; HTGS_PHASE1.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 45545)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burkett,C., Burrell,K.L., Byrd,N.C., Carron,F.P.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgat,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguri,M., Okuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Stinson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalongo,C.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 45545)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627165.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCQV
Center clone name: CH230-8M11
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 20630 bases at least Q40
Consensus quality: 26564 bases at least Q30
Consensus quality: 31670 bases at least Q20
Estimated insert size: 18921; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3472: contig of 3472 bp in length
* 3572: gap of unknown length
* 6120: contig of 2548 bp in length
* 6220: gap of unknown length
* 8187: contig of 1967 bp in length
* 8287: gap of unknown length
* 8188: gap of unknown length
* 10579: contig of 2292 bp in length
* 10679: gap of unknown length
* 13158: contig of 2479 bp in length
* 13258: gap of unknown length
* 13159: contig of 1844 bp in length
* 15102: contig of 1844 bp in length
* 15202: gap of unknown length
* 15203: contig of 1862 bp in length
* 17064: contig of 1862 bp in length
* 17164: gap of unknown length
* 19153: contig of 1987 bp in length
* 19152: gap of unknown length
* 19152: contig of 1357 bp in length
* 20608: gap of unknown length
* 20609: contig of 1168 bp in length
* 21877: gap of unknown length
* 21876: contig of 1168 bp in length
* 21977: contig of 1719 bp in length
* 23696: gap of unknown length
* 23795: gap of unknown length
* 25013: contig of 1218 bp in length
* 25014: gap of unknown length
* 25113: gap of unknown length
* 26739: contig of 1626 bp in length
* 26839: gap of unknown length
* 26840: contig of 1873 bp in length
* 28712: contig of 1873 bp in length
* 28812: gap of unknown length
* 30284: contig of 1472 bp in length
* 30384: gap of unknown length
* 30385: contig of 1209 bp in length
* 31593: gap of unknown length
* 31594: gap of unknown length
* 31694: contig of 1352 bp in length
* 33046: gap of unknown length
* 33146: contig of 1501 bp in length
* 34646: gap of unknown length
* 34747: gap of unknown length
* 36406: contig of 1660 bp in length
* 36506: gap of unknown length
* 36507: contig of 1288 bp in length
* 37795: gap of unknown length
* 37895: contig of 1660 bp in length
* 39555: gap of unknown length
* 39655: contig of 1135 bp in length
* 40790: gap of unknown length
* 40890: contig of 1527 bp in length
* 42417: gap of unknown length
* 42517: contig of 1809 bp in length
* 44326: gap of unknown length
* 44426: contig of 1120 bp in length.
*
* Location/Qualifiers
* 1..45545
* /organism="Rattus norvegicus"
* /db_xref="taxon:10116"
* /clone="CH230-8M11"
*
* BASE COUNT 12026 a 11519 c 9167 g 10397 t 2436 others
* ORIGIN
*
* Query Match 80.0%; Score 19.2; DB 2; Length 45545;
* Best Local Similarity 87.5%; Pred. No. 68;
* Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 tgcctgactttcaactgtgttc 24
||||| || |||||
Db 44495 TGCCCTGTCTGTCTACTCTGTCTC 44518

Search completed: August 24, 2002, 21:54:34
Job time: 15559 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 22:11:57 ; Search time 905.65 Seconds
(without alignments)
45.499 Million cell updates/sec

Title: US-09-986-381-5
Perfect score: 24
Sequence: 1 tgcctgacttcaactctgtctc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	28	AA332239	Primer for amplify
2	24	100.0	28	AA332239	Human p53 gene PCR
3	24	100.0	600	AA332239	Human p53 gene fra
4	22	91.7	121	AA332239	p53 mutation corre
5	22	91.7	121	AA332239	p53 mutation corre
6	19	79.2	23	AA332239	DNA synthesis meth
7	18.8	78.3	6511	AA332239	Drosophila melanog
8	18.8	78.3	11239	AA332239	Drosophila melanog
9	18.8	78.3	11950	AA332239	Human genomic DNA

c 10	18.8	78.3	11950	22	AA332239	Genomic sequence #
c 11	18.8	78.3	11950	22	AA332239	Human endocrine po
c 12	17.8	74.2	965	22	AA332239	Human olfactory re
c 13	17.8	74.2	1024	21	AA332239	Human prostate can
c 14	17.8	74.2	1024	21	AA332239	Human prostate can
c 15	17.8	74.2	1949	22	AA332239	Human cDNA sequenc
c 16	17.8	74.2	3276	22	AA332239	Human nervous syst
c 17	17.6	73.3	1277	21	AA332239	cDNA encoding a nu
c 18	17.6	73.3	3127	23	AA332239	DNA encoding novel
c 19	17.6	73.3	3566	23	AA332239	DNA encoding novel
c 20	17.2	71.7	367	21	AA332239	Human secreted pro
c 21	17.2	71.7	875	22	AA332239	cDNA encoding nove
c 22	17.2	71.7	2040	23	AA332239	Drosophila melanog
c 23	17.2	71.7	2398	22	AA332239	Human immune/haema
c 24	17.2	71.7	4319	23	AA332239	Drosophila melanog
c 25	17.2	71.7	8948	22	AA332239	Human immune/haema
c 26	17.2	71.7	10069	23	AA332239	Drosophila melanog
c 27	17.2	71.7	10812	22	AA332239	DNA encoding novel
c 28	17.2	71.7	10812	22	AA332239	Human reproductive
c 29	17.2	70.8	652	22	AA332239	Novel cDNA encodin
c 30	16.6	69.2	47	23	AA332239	CNS disorder-relat
c 31	16.6	69.2	78	22	AA332239	Human brain expres
c 32	16.6	69.2	189	22	AA332239	Human breast cell
c 33	16.6	69.2	189	22	AA332239	Human foetal liver
c 34	16.6	69.2	189	22	AA332239	Human bone marrow
c 35	16.6	69.2	189	22	AA332239	Probe #9763 for ge
c 36	16.6	69.2	189	22	AA332239	Probe #13711 used
c 37	16.6	69.2	189	22	AA332239	Probe #5537 used t
c 38	16.6	69.2	333	22	AA332239	Human polynucleoti
c 39	16.6	69.2	453	22	AA332239	Human foetal liver
c 40	16.6	69.2	453	22	AA332239	Human bone marrow
c 41	16.6	69.2	453	22	AA332239	Probe #512 for gen
c 42	16.6	69.2	453	22	AA332239	Probe #518 used to
c 43	16.6	69.2	453	22	AA332239	Probe #502 used to
c 44	16.6	69.2	510	22	AA332239	Human brain expres
c 45	16.6	69.2	589	23	AA332239	CNS disorder-relat

ALIGNMENTS

RESULT 1
AA332239 standard; DNA; 28 BP.
XX AC AA332239;
XX DT 14-JUN-1999 (first entry)
XX DE Primer for amplifying DNA containing a p53 mutation.
XX KW Allele profile; diagnosis; treatment; pharmacogenetic; breast cancer;
XX KW CTR; cystic fibrosis; dystrophin; Duchenne muscular dystrophy; p53;
XX KW Becker muscular dystrophy; Li-Fraumeni syndrome; neurofibromatosis;
XX KW colorectal cancer; MSH2 gene; MLH1 gene; BRCA1 gene; BRCA2 gene;
XX KW BAP1 gene; human papillomavirus; HPV; mutation; PCR primer; ss.
XX OS Synthetic.
XX PN WO9906598-A2.
XX PD 11-FEB-1999.
XX PF 04-AUG-1998; 98WO-US16574.
XX PR 22-MAY-1998; 98US-0084471.
XX PR 04-AUG-1997; 97US-0905772.
XX PA (ONCO-) ONCORMED INC.
XX PI Murphy PD;
XX PR 1999-153820/13.
XX DR WPI; 1999-153820/13.

XX
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Determining common functional alleles in a population - useful in the diagnosis of disease associated with allelic heterogeneity

Example 6; Page 46; 78pp; English.

The invention relates to methods of determining a functional allele profile of a gene in a population. Functional allele profiles comprise the commonly occurring alleles in a population, and the relative frequencies at which such alleles of a given gene occur. The methods are used to identify and determine the frequency of the functional alleles of genes which display extensive allelic heterogeneity, particularly those implicated in disease or conditions, such as the BRCA1 gene associated with breast cancer, CFTR associated with cystic fibrosis, dystrophin associated with Duchenne muscular dystrophy and Becker muscular dystrophy, and p53 associated with Li-Fraumeni syndrome. The methods can also be employed for diseases where allelic and genetic heterogeneity exist, such as breast cancer, neurofibromatosis, and hereditary non-polyposis colorectal cancer. Identification of functional alleles is necessary for identification of mutations which may be implicated in the disease. Sequences AAX32233-254 represent primers for sequencing genomic DNA containing a p53 mutation associated with human papillomavirus (HPV) infection. This is used in the pharmacogenetic analysis of a p53 polymorphism.

Sequence 28 BP; 3 A; 10 C; 4 G; 11 T; 0 other;

XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

A method has been developed for performing multiple PCRs in a single vessel. The method comprises priming DNA synthesis on a template in a vessel with at least 2 sets of primers, where the primers are present in a vessel at a predetermined ratio, where the ratio is:
Ca = Cl(La/L1)2; Ca = the concentration of primers for an amplicon a;
Cl = the concentration of primer for the longest amplicon; La = the length of the amplicon a; and L1 is the length of the longest amplicon. Also described is a method of performing multiple PCRs in a single vessel comprising: (a) priming DNA synthesis on a genomic p53 template in a vessel with 10 sets of primers which amplify exons 2-11 of p53, where the primers are shown in AAX18649 to AAX18668. The methods are used to amplify nucleic acids for e.g. paternity testing, forensic analysis, genetic screening, polymorphism detection, and mutation analyses. They are used particularly for amplifying coding regions of the human p53 gene. The methods permit the amplification of different sequences with the same efficiency so that approximately equimolar products result.

Sequence 28 BP; 3 A; 10 C; 4 G; 11 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgccctgactttcaactctgtctc 24
Db 5 tgccctgactttcaactctgtctc 28

RESULT 3
AAT59674
ID AAT59674 standard; DNA; 600 BP.
XX
AC AAT59674;
XX
DT 07-OCT-1997 (first entry)
XX
DE Human p53 gene fragment.
XX
KW Target nucleotide sequence; point mutation; analysis;
KW criminal investigation; parental determination; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 105..288
FT /*tag= a
FT /number= 1
FT /note= "Exon in which possible mutation sites occur"
FT misc_difference 163
FT /*tag= b
FT /note= "Wild-type T could be point mutated to G"
FT misc_difference 267
FT /*tag= c
FT /note= "Wild-type G could be point mutated to C"
XX
PN WO9641001-A1.
XX
XX
PD 19-DEC-1996.
XX
XX 05-JUN-1996; 96WO-US08759.
XX
XX 07-JUN-1995; 95US-0472239.
XX
XX (ONCO-) ONCOR INC.
XX
XX Bhatnagar SK, George AL, Nazarenko I;
XX WPI: 1997-099941/09.
XX
XX Analysis of target nucleotide sequence which exits in a first state
PT

XX
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Determining common functional alleles in a population - useful in the diagnosis of disease associated with allelic heterogeneity

Example 6; Page 46; 78pp; English.

The invention relates to methods of determining a functional allele profile of a gene in a population. Functional allele profiles comprise the commonly occurring alleles in a population, and the relative frequencies at which such alleles of a given gene occur. The methods are used to identify and determine the frequency of the functional alleles of genes which display extensive allelic heterogeneity, particularly those implicated in disease or conditions, such as the BRCA1 gene associated with breast cancer, CFTR associated with cystic fibrosis, dystrophin associated with Duchenne muscular dystrophy and Becker muscular dystrophy, and p53 associated with Li-Fraumeni syndrome. The methods can also be employed for diseases where allelic and genetic heterogeneity exist, such as breast cancer, neurofibromatosis, and hereditary non-polyposis colorectal cancer. Identification of functional alleles is necessary for identification of mutations which may be implicated in the disease. Sequences AAX32233-254 represent primers for sequencing genomic DNA containing a p53 mutation associated with human papillomavirus (HPV) infection. This is used in the pharmacogenetic analysis of a p53 polymorphism.

Sequence 28 BP; 3 A; 10 C; 4 G; 11 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgccctgactttcaactctgtctc 24
Db 5 tgccctgactttcaactctgtctc 28

RESULT 2
AAX18655
ID AAX18655 standard; DNA; 28 BP.
XX
AC AAX18655;
XX
DT 06-MAY-1999 (first entry)
XX
DE Human p53 gene PCR primer #7.
XX
KW Human p53; multiplex amplification; single reaction vessel;
KW paternity testing; forensic analysis; genetic screening;
KW polymorphism detection; mutation analysis; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9859072-A1.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-US12779.
XX
XX 20-JUN-1997; 97US-0050405.
XX
XX (AFFY-) AFFYMETRIX INC.
XX
XX Matsuzaki H, Murphy EA;
XX
XX WPI: 1999-105631/09.
XX
XX Carrying out multiple PCRs in a single vessel - by priming DNA
PT synthesis using primers in a predetermined ratio to produce
PT equimolar amounts of the different products
XX
XX Claim 7; Page 12; 25pp; English.

Figure D

PT or different second state - useful for determination of point
PT mutation(s)
XX
PS Disclosure; Page 46; 78pp; English.
XX
CC A novel method has been produced for analysing a target nucleotide
CC sequence. The target sequence having a first segment, a second segment
CC and a third segment between the first and second, which is formed of at
CC least one, but less than four different, nucleotide(s) where the third
CC segment has a nucleotide or nucleotide sequence in a first state or
CC different second state. The present sequence is a partial human p53
CC gene sequence which is used as an example of a target nucleotide
CC sequence. The method can be used to analyse the entire sequence
CC (portion) of a known gene and to analyse infectious disease or to
CC determine if a sample is from a particular source, such as for criminal
CC investigations or parental determination. Especially the method is used
CC to determine the existence or absence of mutations consisting of one or
CC a few nucleotides.
XX
SQ Sequence 600 BP; 102 A; 170 C; 168 G; 160 T; 0 other;

Query Match 100.0%; Score 24; DB 18; Length 600;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcctgacttcaactctgtctc 24
|||||

Db 65 tgcctgacttcaactctgtctc 88

RESULT 4

ABA77347
ID ABA77347 standard; DNA; 121 BP.

XX ABA77347;

XX 24-JAN-2002 (first entry)

XX p53 mutation correcting oligonucleotide SEQ ID NO: 193.

XX Human; gene therapy; adenosine deaminase deficiency; p53; beta-globin;
KW retinoblastoma; BRCA1; BRCA2; CFTR; cystic fibrosis; cancer; Factor V;
KW cyclin-dependent kinase inhibitor 2A; CDKN2A; melanoma; APC; HBA1; HBA2;
KW adenomatous polyposis of the colon; Factor VII; Factor IX; thrombosis;
KW haemophilia; alpha thalassaemia; haemoglobin alpha locus 1; MLH1; APOE;
KW mismatch repair; MSH2; MSH6; hyperlipidaemia; apolipoprotein E; LDLR;
KW familial hypercholesterolaemia; UGT1; syndrome; APP; PSEN1; antisense;
KW UDP-glucuronosyltransferase; amyloid precursor protein; presenilin-1;
KW Alzheimer's disease; cytostatic; antiskilling; antianaemic; haemostatic;
KW antilipemic; ss.

XX Homo sapiens.

XX WO200173002-A2.

XX 04-OCT-2001.

XX 27-MAR-2001; 2001WO-US09761.

XX 27-MAR-2000; 2000US-192176P.

XX 27-MAR-2000; 2000US-192179P.

XX 01-JUN-2000; 2000US-208538P.

XX 30-OCT-2000; 2000US-244989P.

XX (UYDE) UNIV DELAWARE.

XX Kmiec EB, Gamper HB, Rice MC;

XX WPI; 2001-639230/73.

XX Oligonucleotide for targeted alterations of genetic sequences and for

PT treating cystic fibrosis, comprises at least one mismatch and chemical

PT modification -
XX Claim 7; Page 54; 294pp; English.
XX
CC The present invention provides single-stranded oligonucleotides which can
CC be used for the targeted alteration of genomic sequences, where the
CC oligonucleotide has at least one mismatch compared with the genomic
CC sequence to be altered. In particular, these sequences are directed at
CC the following genes: adenosine deaminase, p53, beta-globin,
CC retinoblastoma, BRCA1, BRCA2, CFTR, cyclin-dependent kinase inhibitor 2A
CC (CDKN2A), APC, Factor V, Factor VIII, Factor IX, haemoglobin alpha locus
CC 1 (HBA1), haemoglobin alpha locus 2 (HBA2), MLH1, MSH2, MSH6,
CC apolipoprotein E (APOE), LDL receptor (LDLR), UDP-glucuronosyltransferase
CC (UGT1), amyloid precursor protein (APP), presenilin-1 (PSEN1) and
CC presenilin-2 (PSEN2). These can be used in the gene therapy of diseases
CC such as cancer, adenosine deaminase deficiency, cystic fibrosis,
CC haemophilia, hypercholesterolaemia, thalassaemia, sickle cell anaemia,
CC Alzheimer's disease, melanoma, adenomatous polyposis of the colon and
CC various syndromes. The present sequence is one of the gene correcting
CC oligonucleotides of the invention.
XX
SQ Sequence 121 BP; 20 A; 48 C; 20 G; 33 T; 0 other;

Query Match 91.7%; Score 22; DB 22; Length 121;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cccgtgacttcaactctgtctc 24

|||||

Db 1 cccgtgacttcaactctgtctc 22

RESULT 5

ABA77348/C

ID ABA77348 standard; DNA; 121 BP.

XX ABA77348;

XX 24-JAN-2002 (first entry)

XX p53 mutation correcting oligonucleotide SEQ ID NO: 194.

XX Human; gene therapy; adenosine deaminase deficiency; p53; beta-globin;
KW retinoblastoma; BRCA1; BRCA2; CFTR; cystic fibrosis; cancer; Factor V;
KW cyclin-dependent kinase inhibitor 2A; CDKN2A; melanoma; APC; HBA1; HBA2;
KW adenomatous polyposis of the colon; Factor VII; Factor IX; thrombosis;
KW haemophilia; alpha thalassaemia; haemoglobin alpha locus 1; MLH1; APOE;
KW mismatch repair; MSH2; MSH6; hyperlipidaemia; apolipoprotein E; LDLR;
KW familial hypercholesterolaemia; UGT1; syndrome; APP; PSEN1; antisense;
KW UDP-glucuronosyltransferase; amyloid precursor protein; presenilin-1;
KW Alzheimer's disease; cytostatic; antiskilling; antianaemic; haemostatic;
KW antilipemic; ss.

XX Homo sapiens.

XX WO200173002-A2.

XX 04-OCT-2001.

XX 27-MAR-2001; 2001WO-US09761.

XX 27-MAR-2000; 2000US-192176P.

XX 27-MAR-2000; 2000US-192179P.

XX 01-JUN-2000; 2000US-208538P.

XX 30-OCT-2000; 2000US-244989P.

XX (UYDE) UNIV DELAWARE.

XX Kmiec EB, Gamper HB, Rice MC;

XX WPI; 2001-639230/73.

XX

PT Oligonucleotide for targeted alterations of genetic sequences and for
PT treating cystic fibrosis, comprises at least one mismatch and chemical
PT modification -
XX
PS Claim 7; Page 54; 294pp; English.
XX
XX The present invention provides single-stranded oligonucleotides which can
CC be used for the targeted alteration of genomic sequences, where the
CC oligonucleotide has at least one mismatch compared with the genomic
CC sequence to be altered. In particular, these sequences are directed at
CC the following genes: adenosine deaminase, p53, beta-globin,
CC retinoblastoma, BRCA1, BRCA2, CTRF, cyclin-dependent kinase inhibitor 2A
CC (CDKN2A), APC, Factor V, Factor VIII, Factor IX, haemoglobin alpha locus
CC 1 (HBA1), haemoglobin alpha locus 2 (HBA2), MLH1, MSH2, MSH6,
CC apolipoprotein E (APOE), LDL receptor (LDLR), UDP-glucuronosyltransferase
CC (UGT1), amyloid precursor protein (APC), presenilin-1 (PSEN1) and
CC presenilin-2 (PSEN2). These can be used in the gene therapy of diseases
CC such as cancer, adenosine deaminase deficiency, cystic fibrosis,
CC haemophilia, hypercholesterolaemia, thalassaemia, sickle cell anaemia,
CC Alzheimer's disease, melanoma, adenomatous polyposis of the colon and
CC various syndromes. The present sequence is one of the gene correcting
CC oligonucleotides of the invention.
XX
SQ Sequence 121 BP; 33 A; 20 C; 48 G; 20 T; 0 other;

Query Match 91.7%; Score 22; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cccctgacttcaactctgtctc 24
|||||
Db 121 CCCTGACTTTCACACTGTCTC 100

RESULT 6
AAF85579
ID AAF85579 standard; DNA: 23 BP.
XX
XX AAF85579;
AC
XX
XX 25-JUN-2001 (first entry)
XX
XX DNA synthesis method PCR primer 13011fw62.
XX
XX DNA synthesis; enzyme activation; sequence determination; diagnosis;
KW forensic medicine; cohort genetics; PCR primer; ss.
KW
XX Unidentified.
XX
XX JP2001017192-A.
XX
XX 23-JAN-2001.
XX
XX 26-JUN-2000; 2000JP-0191940.
XX
XX 24-JUN-1999; 99US-0339103.
XX
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
XX WPI; 2001-285411/30.
XX
XX Synthesis of nucleic acids useful for useful for sequence determination
PT in diagnosis, forensic medicine and cohort genetics comprises
PT continuous activation of enzyme(s) in heat cycle reaction -
XX
XX Example 3; Page 15; 22pp; Japanese.
XX

XX The present invention describes a method for the synthesis of a nucleic
CC acid molecules, involving the continuous activation of enzyme(s) in heat
CC cycle reaction. This is useful in the synthesis of nucleic acids,
CC especially DNA molecular synthesis, sequence determination in diagnosis,
CC forensic medicine and cohort genetics. The present sequence is a PCR

CC primer used to demonstrate the method.
XX
SQ Sequence 23 BP; 3 A; 8 C; 3 G; 9 T; 0 other;

Query Match 79.2%; Score 19; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgccctgacttcaactct 19
|||||
Db 5 tgccctgacttcaactct 23

RESULT 7
ABL12613/c
ID ABL12613 standard; cDNA; 6511 BP.
XX

XX ABL12613;
AC

XX 26-MAR-2002 (first entry)
DT

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 32321.
DE

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
KW

XX Drosophila melanogaster.
OS

XX WO200171042-A2.
PN

XX 27-SEP-2001.
PD

XX 23-MAR-2001; 2001WO-US09231.
PF

XX 23-MAR-2000; 2000US-191637P.
PR

XX 11-JUL-2000; 2000US-0614150.
PR

XX (PEKE) PE CORP NY.
PA

XX Venter JC, Adams M, Li PWD, Myers EW;
PI

XX WPI; 2001-656860/75.
DR

XX P-PSDB; ABB68510.
DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT

XX Claim 1; SEQ ID NO 32321; 2lpp + Sequence Listing; English.
PS

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 6511 BP; 1951 A; 1514 C; 1706 G; 1340 T; 0 other;

Query Match 78.3%; Score 18.8; DB 23; Length 6511;
Best Local Similarity 90.9%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgccctgacttcaactctgtc 22
|||||
Db 5462 TGCCCTGACTATCAACTCTGAC 5441


```
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 06-DEC-2000; 2000US-256719P.
PR 08-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.

PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476161/51.
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX Disclosure; SEQ ID NO: 899; 859pp + Sequence Listing; English.
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a genomic DNA of the invention.
XX Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 other;

Query Match 78.3%; Score 18.8; DB 22; Length 11950;
Best Local Similarity 90.9%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgcctgactttcaactctgtc 22
   |||||
Db 326 TGCCTGACTTTCCTCTGTC 305

RESULT 10
AAS41753/c
ID AAS41753 standard; DNA; 11950 BP.
XX AAS41753;
XX AC
XX DT
XX DE Genomic sequence #69 encoding novel human enzyme polypeptide.
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; ds.
XX Homo sapiens.
XX OS
XX PN WO200155301-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01239.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
```

PR	11-JUL-2000;	2000US-0217496;
PR	14-JUL-2000;	2000US-0218290;
PR	26-JUL-2000;	2000US-0220963;
PR	26-JUL-2000;	2000US-0220964;
PR	14-AUG-2000;	2000US-0224518;
PR	14-AUG-2000;	2000US-0224519;
PR	14-AUG-2000;	2000US-0225213;
PR	14-AUG-2000;	2000US-0225214;
PR	14-AUG-2000;	2000US-0225266;
PR	14-AUG-2000;	2000US-0225267;
PR	14-AUG-2000;	2000US-0225268;
PR	14-AUG-2000;	2000US-0225270;
PR	14-AUG-2000;	2000US-0225447;
PR	14-AUG-2000;	2000US-0225757;
PR	14-AUG-2000;	2000US-0225758;
PR	14-AUG-2000;	2000US-0225759;
PR	18-AUG-2000;	2000US-0226279;
PR	22-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226868;
PR	22-AUG-2000;	2000US-0227182;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-0229287;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	05-SEP-2000;	2000US-0229509;
PR	06-SEP-2000;	2000US-0229513;
PR	06-SEP-2000;	2000US-0230437;
PR	06-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0231242;
PR	08-SEP-2000;	2000US-0231243;
PR	08-SEP-2000;	2000US-0231244;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0232080;
PR	12-SEP-2000;	2000US-0232081;
PR	14-SEP-2000;	2000US-02321968;
PR	14-SEP-2000;	2000US-0232397;
PR	14-SEP-2000;	2000US-0232398;
PR	14-SEP-2000;	2000US-0232399;
PR	14-SEP-2000;	2000US-0232400;
PR	14-SEP-2000;	2000US-0232401;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234223;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0235484;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	27-SEP-2000;	2000US-0236227;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	13-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241221;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;

PR	01-NOV-2000;	2000US-0246174.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0256179.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251969.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
FA	(HUMA-) HUMAN GENOME SCI INC	
XX		
PI	Rosen CA, Barash SC, Ruben	
XX		
DR	WPI; 2001-465566/50.	
XX		
PT	Novel polypeptides and polymers	
PT	preventing, treating neural, and	
PT	pulmonary, cardiovascular, renal	
PT	diseases -	
XX		
XX		
PS	Disclosure; SEQ ID NO 1879;	
XX		
CC	The present invention relates to	
CC	polypeptides (AAU22915-AAU233	
CC	encoding them. The enzyme poly	
CC	functional classes of oxidoreduc	
CC	isomerases or ligases. The sub	
CC	diagnosis, treatment, prevention	
CC	disorders including hyperprolifer	
CC	immunodeficiency disorders (e.g.	
CC	arthritis), neurologic disorders	
CC	metabolic disorders (e.g. phos	

CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS41685-AAS42192 represent DNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 other;

Query Match 78.3%; Score 18.8; DB 22; Length 11950;
Best Local Similarity 90.9%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 tgcctgacttcaactctgac 22
|||||
Db 326 TGCCCTGACTTCTCCTCTGTC 305

RESULT 11
AAS29743/C
ID AAS29743 standard; DNA; 11950 BP.

AC AAS29743;

XX 21-NOV-2001 (first entry)

XX Human endocrine polypeptide encoding genomic DNA SEQ ID NO 469.

XX Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility.

XX Homo sapiens.

XX W020015364-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01308.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-451936/48.

XX Isolated polypeptide for treating, preventing and/or

PT disorders of the endocrine system such as reproductive disorders,

PT endocrine cancers and also for testing and detection e.g. diagnosis -

XX Claim 1; SEQ ID No 469; 604pp; English.

XX Sequences AAS29737-AAS29769 represent genomic DNA molecules, which encode
CC the endocrine polypeptides of the invention. Endocrine polypeptides and
CC their associated polynucleotides of the invention are useful in the
CC diagnosis, treatment and prevention of various types of disorders in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.

CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match 78.3%; Score 18.8; DB 22; Length 11950;
Best Local Similarity 90.9%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgcctgactttcaactctgtc 22
|||||
Db 326 TGCCCTGACTTTCTCTCTGTC 305

RESULT 12

AAH31581
ID AAH31581 standard; DNA; 965 BP.

XX AAH31581;

XX 30-JUL-2001 (first entry)

XX Human olfactory receptor polynucleotide, SEQ ID NO: 154.

XX Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.

XX Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX Claim 8; Page 236; 1857pp; English.

XX The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.

XX Sequence 965 BP; 196 A; 273 C; 198 G; 298 T; 0 other;

Query Match 74.2%; Score 17.8; DB 22; Length 965;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PP 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 XX WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 17869; 2537pp + CD ROM; English.
 PS
 XX
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1949 BP; 520 A; 453 C; 379 G; 597 T; 0 other;

Query Match 74.2%; Score 17.8; DB 22; Length 1949;
 Best Local Similarity 90.5%; Pred. No. 1.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 cctgaacttcaactctgtctc 24
 Db 1390 cttgactttcaactctgtctc 1410

Search completed: August 24, 2002, 22:12:00
 Job time: 11970 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:55:44 ; Search time 204.64 seconds
(without alignments)
28.808 Million cell updates/sec

Title: US-09-986-381-5
Perfect score: 24
Sequence: 1 tgcctgactttcaactctgtctc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	24	100.0	28	4	US-09-099-301-7
2	24	100.0	600	1	US-08-472-239-2
3	16.6	69.2	10807	1	US-08-206-176-7
4	16.6	69.2	10807	2	US-08-756-506-5
5	16	66.7	22	2	US-08-441-887A-288
6	16	66.7	49	2	US-08-441-887A-289
7	16	66.7	245	3	US-08-976-255-6
8	16	66.7	309	2	US-08-785-571-6
9	16	66.7	309	5	PCT-US93-06734-6
10	16	66.7	333	2	US-08-785-571-8
11	16	66.7	333	5	PCT-US93-06734-8
12	16	66.7	801	3	US-08-976-255-4
13	16	66.7	1600	2	US-08-933-750C-50
14	16	66.7	1600	3	US-09-234-613-50
15	16	66.7	1622	4	US-09-334-601-9
16	15.8	65.8	1078	3	US-08-981-256A-4
17	15.6	65.0	640	4	US-09-328-111-792
18	15.6	65.0	2220	5	PCT-US95-13749-2
19	15.6	65.0	2443	2	US-08-685-625A-1
20	15.2	63.3	669	4	US-09-079-984A-2
21	15.2	63.3	669	4	US-09-079-984A-12
22	15.2	63.3	1318	2	US-08-453-051-1
23	15.2	63.3	1716	2	US-08-954-333-9
24	15.2	63.3	2421	1	US-08-063-552-1
25	15.2	63.3	2421	5	PCT-US93-05704-1
26	15.2	63.3	4430	2	US-08-918-914-2
27	15.2	63.3	8910	3	US-08-779-764A-1

28	15.2	63.3	10627	1	US-08-060-925A-12	Sequence 12, Appl
29	15	62.5	29	3	US-09-280-799-33	Sequence 33, Appl
30	15	62.5	714	4	US-08-943-731-115	Sequence 115, App
c 31	15	62.5	784	4	US-08-998-416-267	Sequence 267, App
c 32	15	62.5	1184	1	US-08-362-511A-4	Sequence 4, Appli
c 33	15	62.5	1239	1	US-08-362-511A-1	Sequence 1, Appli
c 34	15	62.5	1534	3	US-08-629-643A-4	Sequence 4, Appli
35	15	62.5	1534	4	US-09-155-884-4	Sequence 4, Appli
36	15	62.5	2000	1	US-08-351-473B-1	Sequence 1, Appli
c 37	15	62.5	2158	4	US-09-318-448-4	Sequence 4, Appli
38	15	62.5	2273	2	US-08-788-539A-1	Sequence 1, Appli
c 39	15	62.5	2437	1	US-07-795-859B-5	Sequence 5, Appli
c 40	15	62.5	2437	1	US-08-457-616-5	Sequence 104, App
c 41	15	62.5	3358	2	US-08-469-537A-104	Sequence 1, Appli
42	15	62.5	3503	1	US-07-631-717A-1	Sequence 1, Appli
43	15	62.5	3503	1	US-08-166-717D-1	Sequence 1, Appli
44	15	62.5	4086	1	US-08-313-181-1	Sequence 1, Appli
c 45	15	62.5	5427	1	US-08-168-917-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-099-301-7
; Sequence 7, Application US/09099301
; Patent No. 6333179
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; TITLE OF INVENTION: Multiplex Amplification of Nucleic Acids
; FILE REFERENCE: 03848.74891
; CURRENT APPLICATION NUMBER: US/09/099,301
; CURRENT FILING DATE: 1998-06-18
; EARLIER APPLICATION NUMBER: 60/050,405
; EARLIER FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: coding sequence of human p53 gene.
US-09-099-301-7

Query Match 100.0%; Score 24; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.0047; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

QY 1 tgcctgactttcaactctgtctc 24
|||||
Db 5 tgcctgactttcaactctgtctc 28

RESULT 2
US-08-472-239-2
; Sequence 2, Application US/08472239
; Patent No. 5728526
; GENERAL INFORMATION:
; APPLICANT: GEORGE, Jr., Albert L.
; APPLICANT: BHATNAGAR, Satish K.
; APPLICANT: NAZARENKO, Irena
; TITLE OF INVENTION: METHOD FOR ANALYZING A NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE AND KIT THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States

```
;
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,239
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mukai, Robert G.
; REGISTRATION NUMBER: 28,531
; REFERENCE/DOCKET NUMBER: 020160-229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-472-239-2

Query Match 100.0%; Score 24; DB 1; Length 600;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgccctgactttcaactgtgtctc 24
    ||||| ||||| ||||| ||||| |||||
Db 65 TGCCCTGACTTCAACTGTGTCTC 88

RESULT 3
US-08-206-176-7/G
; Sequence 7, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A
; APPLICANT: Prunkard, Donna E
; APPLICANT: Foster, Donald C
; TITLE OF INVENTION: Production of Fibrinogen in Transgenic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-15
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-547-8080 ext 322
; TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ovine beta-lactoglobulin
; US-08-206-176-7

Query Match 69.2%; Score 16.6; DB 1; Length 10807;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gccctgactttcaactgtgtctc 24
    ||||| ||||| ||||| ||||| ||||| ||
Db 6980 GCCCTCACTGTCAACTCTGGATC 6958

RESULT 4
US-08-756-506-5/C
; Sequence 5, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,506
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 95-28
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-756-506-5
```

```
Query Match 69.2%; Score 16.6; DB 2; Length 10807;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gccctgactttcaactgtgtctc 24
    ||||| ||||| ||||| ||||| ||||| ||
Db 6980 GCCCTCACTGTCAACTCTGGATC 6958
```

RESULT 5
US-08-441-887A-288
; Sequence 288, Application US/08441887A
; Patent No. 5837832
; GENERAL INFORMATION:
; APPLICANT: Chee, Mark
; APPLICANT: Cronin, Maureen T.
; APPLICANT: Fodor, Stephen P.A.
; APPLICANT: Huang, Xiaohua X.
; APPLICANT: Hubbell, Earl A.
; APPLICANT: Lipschutz, Robert J.
; APPLICANT: Lobban, Peter E.
; APPLICANT: Morris, Macdonald S.
; APPLICANT: Sheldon, Edward L.
; TITLE OF INVENTION: Arrays of Nucleic Acid Probes on
; TITLE OF INVENTION: Biological Chips
; NUMBER OF SEQUENCES: 360
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441.887A
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/143,312
; FILING DATE: 26-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/082,937
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joseph O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 018547-00416005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-326-2400
; TELEFAX: 650-326-2422
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (probe)
US-08-441-887A-288

Query Match 66.7%; Score 16; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgccttgactttcaac 16
| | | | | | | | | | | | | | | | | | | | | |
Db 7 TGCCTGACTTTCAAC 22

RESULT 6
US-08-441-887A-289
; Sequence 289, Application US/08441887A
; Patent No. 5837832
; GENERAL INFORMATION:
; APPLICANT: Chee, Mark
; APPLICANT: Cronin, Maureen T.

; APPLICANT: Fodor, Stephen P.A.
; APPLICANT: Huang, Xiaohua X.
; APPLICANT: Hubbell, Earl A.
; APPLICANT: Lipschutz, Robert J.
; APPLICANT: Lobban, Peter E.
; APPLICANT: Morris, Macdonald S.
; APPLICANT: Sheldon, Edward L.
; TITLE OF INVENTION: Arrays of Nucleic Acid Probes on
; TITLE OF INVENTION: Biological Chips
; NUMBER OF SEQUENCES: 360
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441.887A
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/143,312
; FILING DATE: 26-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/082,937
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joseph O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 018547-00416005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-326-2400
; TELEFAX: 650-326-2422
; INFORMATION FOR SEQ ID NO: 289:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (probe)
US-08-441-887A-289

Query Match 66.7%; Score 16; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgccttgactttcaac 16
| | | | | | | | | | | | | | | | | | | | | |
Db 34 TGCCTGACTTTCAAC 49

RESULT 7
US-08-976-255-6/c
; Sequence 6, Application US/08976255
; Patent No. 6136581
; GENERAL INFORMATION:
; APPLICANT: Jono, Keith E.
; APPLICANT: Plowman, Gregory
; TITLE OF INVENTION: KINASE GENES AND USES
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles

```
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,255
; FILING DATE: NO. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031,675
; FILING DATE: NO. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..309
; US-08-976-255-6

Query Match 66.7%; Score 16; DB 3; Length 245;
Best Local Similarity 79.2%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tgcctgactttcaactctgtctc 24
   ||| ||||| ||||| |||
Db 96 TGCAC TACTTTTAACTCTCTTCAC 73

RESULT 8
US-08-785-571-6/c
; Sequence 6, Application US/08785571
; Patent No. 5854070
; GENERAL INFORMATION:
; APPLICANT: Rose, Lynn M.
; TITLE OF INVENTION: Alleviation of Symptoms Associated with
; TITLE OF INVENTION: Inflammatory Disease States
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,571
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,089
; FILING DATE:
; APPLICATION NUMBER: US/08/094,535
; FILING DATE:
; APPLICATION NUMBER: US 08/060,699

; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,255
; FILING DATE: NO. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031,675
; FILING DATE: NO. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..309
; US-08-976-255-6

Query Match 66.7%; Score 16; DB 2; Length 309;
Best Local Similarity 79.2%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tgcctgactttcaactctgtctc 24
   |||| | ||||| ||||| |||||
Db 76 TGCCATAACTTTTCAACGCTTTCAC 53

RESULT 9
PCT-US93-06734-6/c
; Sequence 6, Application PC/TUS9306734
; GENERAL INFORMATION:
; APPLICANT: Rose, Lynn M.
; TITLE OF INVENTION: Alleviation of Symptoms Associated with
; TITLE OF INVENTION: Inflammatory Disease States
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06734
; FILING DATE: 19930716
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/060,699
; FILING DATE: 10-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/915,068
; FILING DATE: 16-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25,447
; REFERENCE/DOCKET NUMBER: 31574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)474-6300
; TELEFAX: (312)474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
```


; NAME/KEY: C

Qy 1 tgcctgactttcaactctgtctc 24
 ||||| ||||| ||||| |||||
 Db 100 TGCCATAACTTTCAACGCTTTCAC 77

RESULT 12
US-08-976-255-4/c
; Sequence 4, Application US/08976255
; Patent No. 6136581
; GENERAL INFORMATION:
; APPLICANT: Jono, Keith E.
; APPLICANT: Ploman, Gregory
; TITLE OF INVENTION: KINASE GENES AND USES
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,255
; FILING DATE: No. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031,675
; FILING DATE: No. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-976-255-4

Query Match 66.7%; Score 16; DB 3; Length 801;
Best Local Similarity 79.2%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tgccctgactttcaactctgtctc 24
||| ||||| |||| |||
Db 306 TGCACGACTTTTAACTCTCTAC 283

RESULT 13
US-08-933-750C-50/c
; Sequence 50, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: U937NOT01
; CLONE: 133
; US-08-933-750C-50

Query Match 66.7%; Score 16; DB 2; Length 1600;
Best Local Similarity 79.2%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tgccctgactttcaactctgtctc 24
||| ||||| |||| |||
Db 156 TGCCATAACTTTGAAGTCTCTTC 133

RESULT 14
US-09-234-613-50/c
; Sequence 50, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0

Mon Aug 26 09:57:56 2002

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/234,613
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/933,750
;; FILING DATE: September 23, 1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0356 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 50:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1600 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: U937NOT01
;; CLONE: 133
;; US-09-234-613-50

Query Match 56.7%; Score 16; DB 3; Length 1600;
Best Local Similarity 79.2%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tgcctgactttcaactctgtctc 24
||||| ||||| ||||| ||
Db 156 TGCATAACTTTGAAGTCTGTTTC 133

RESULT 15
US-09-334-601-9
;; Sequence 9, Application US/09334601
;; Patent No. 6280989
;; GENERAL INFORMATION:
;; APPLICANT: Kapitonov, Dmitri
;; TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;; FILE REFERENCE: VCUIP-6
;; CURRENT APPLICATION NUMBER: US/09/334,601
;; CURRENT FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 94
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 9
;; LENGTH: 1622
;; TYPE: DNA
;; ORGANISM: Murinae gen. sp.
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (498)..(1484)
US-09-334-601-9

Query Match 66.7%; Score 16; DB 4; Length 1622;
Best Local Similarity 79.2%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tgcctgactttcaactctgtctc 24
||||| ||||| ||||| ||
Db 461 tccctcactttggactctgttc 484

Search completed: August 24, 2002, 21:55:46
Job time: 13141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:08:21 ; Search time 7654.23 Seconds
(without alignments)
42.320 Million cell updates/sec

Title: US-09-986-381-5
Perfect score: 24
Sequence: 1 tgcctgactttcaactctgtctc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estm:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vit:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	80.8	573	10	BI442839
2	19.4	80.8	634	10	BJ070145
3	19.2	80.0	313	9	BB500878
c 4	19.2	80.0	391	9	AW703308
5	19.2	80.0	488	12	AQ385199
c 6	19.2	80.0	636	9	A1415161
7	19.2	80.0	643	12	AG089634
8	19	79.2	308	9	AV046577
c 9	18.8	78.3	486	10	BG814329
10	18.8	78.3	798	12	AQ329245
11	18.4	76.7	375	9	BB842527
12	18.4	76.7	443	9	BB860133
13	18.4	76.7	453	10	BE553586
14	18.4	76.7	454	9	AW211667
15	18.4	76.7	509	9	BB866609
16	18.4	76.7	589	10	BE282156
17	18.4	76.7	616	9	BB619323

18 18.4 76.7 762 10 BI855252
19 18.4 76.7 870 10 BI167537
20 18.4 76.7 906 10 BI556774
21 18.2 75.8 290 9 BB053033
c 22 18.2 75.8 307 12 AZ079288
c 23 18.2 75.8 362 12 AZ765423
c 24 18.2 75.8 414 9 AA974610
c 25 18.2 75.8 508 12 AQ205376
c 26 18.2 75.8 512 12 AZ226374
c 27 18.2 75.8 548 9 BE031922
c 28 18.2 75.8 565 12 AZ372903
c 29 18.2 75.8 578 12 BH061137
c 30 18.2 75.8 585 10 BG732803
c 31 18.2 75.8 587 12 AZ404092
c 32 18.2 75.8 588 12 AQ496782
c 33 18.2 75.8 624 12 BH050857
c 34 18.2 75.8 664 9 BB221748
c 35 18.2 75.8 672 12 AZ337948
c 36 18.2 75.8 764 10 BI905558
c 37 18.2 75.8 781 10 BG923925
c 38 18.2 75.8 899 10 BE271173
c 39 18.2 75.8 973 12 CNS00KW9
c 40 18.2 75.8 1038 10 BI521674
41 17.8 74.2 265 9 AW920362
42 17.8 74.2 346 10 BG264335
43 17.8 74.2 356 9 AI591885
44 17.8 74.2 382 10 W31685
c 45 17.8 74.2 392 9 AI189949

ALIGNMENTS

RESULT 1
BI442839

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI442839 573 bp mRNA linear EST 21-AUG-2001
dai91a10.y1 NICHD XGC Ovl Xenopus laevis cDNA clone IMAGE:5047602
5' similar to SW:P85A_HUMAN P27986 PHOSPHATIDYLINOSITOL 3-KINASE
REGULATORY ALPHA SUBUNIT ; , mRNA sequence.

BI442839 GI:15267546

EST.

Xenopus laevis

African clawed frog.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 573)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other_ESTs: dai91a10.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

High quality sequence stop: 442.

Location/Qualifiers

1..573

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:5047602"

/clone_lib="NICHD XGC Ovl"

/sex="female"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: PCMV-SPORT6; Site_1: NotI;

Site_2: Sall; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 2.0 kb. Constructed by Life

FEATURES
source


```

Db 48 TGCCCTGACTTTCACCTCTCTC 71
|||||
RESULT 4
AW703308/c
LOCUS
DEFINITION TgESTz292g10.y1 TGRH*-Tachyzoite cDNA Toxoplasma gondii cDNA clone
ACCESSION AW703308
VERSION AW703308.1 GI:7587491
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 391)
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioaka,J.A., Aslett,M.A.,
Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
Waterston,R.H. and Boothroyd,J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
CONTACT: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu
Contact David Sibley (toxoset@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40Rp from Gibco
High quality sequence stop: 350.
FEATURES
source
Location/Qualifiers
1..391
/organism="Toxoplasma gondii"
/strain="RH (Type I)"
/db_xref="taxon:5811"
/clone="TgESTz292g10.y1"
/clone_lib="TGRH*-Tachyzoite cDNA"
/dev_stage="tachyzoite"
/lab_host="SOLR cells"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was synthesized from poly mRNA using an
oligo-dT primer containing an XhoI site. Following second
strand synthesis, EcoRI adapters were ligated to the cDNA
and products were size-selected on sephacryl S500. The
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII
(Stratagene). Clones were converted to phagemids by mass
excision using ExAssist helper phage and E.coli SOLR cells
(Stratagene). Insert sizes range from 0.3-3.0 kb. The
library may contain a small percentage of host or
bacterial contaminants. NOTE: Many clones lack the 5'
EcoRI site but contain the upstream polylinker sites."
BASE COUNT 126 a 78 c 142 g 45 t
ORIGIN
Query Match 80.0%; Score 19.2; DB 9; Length 391;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgccctgacttcaactctgtctc 24
|||||
Db 383 TGCTCTGACTTTCGCTCTGTCTC 360
|||||
RESULT 5
AQ385199
LOCUS
DEFINITION RPC111-141K11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-141K11,
DNA sequence.
ACCESSION AQ385199

```

```

AQ385199.1 GI:4356222
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 488)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other_GSSs: RPC111-141K11.TJ
CONTACT: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..488
/organism="Homo sapiens"
/db_xref="GDB:7554010"
/db_xref="taxon:9606"
/clone="RPCI-11-141K11"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT 101 a 107 c 87 g 192 t 1 others
ORIGIN
Query Match 80.0%; Score 19.2; DB 12; Length 488;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgccctgacttcaactctgtctc 24
|||||
Db 191 TCCCCAGACTTCAACTCTGTCTC 214
|||||
RESULT 6
AI415161/c
LOCUS
DEFINITION IMAGE:332214 3', mRNA sequence.
ACCESSION AI415161
VERSION AI415161.1 GI:4258665
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 636)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999

```

LIBRARY

Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..643
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-088P05.P"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
140 a 162 c 168 g 173 t

80.0%; Score 19.2; DB 12; Length 643;
Similarity 87.5%; Pred. No. 3.1e+02;
1; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ccctgactttcaactctgtctc 24
|||||
TCTGAGTTTCAACTTTGTCTC 414

V0046577 308 bp mRNA linear EST 23-NOV-1999
V0046577 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
7000058H21, mRNA sequence.
V0046577
V0046577.2 GI:4866242

ST.
mouse mouse.
Mus musculus
Mammalia; Chordata; Metazoa; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
(bases 1 to 308)
narninci.P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kuchiki,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomioka,N.,
Tanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
ruramatsu,M., Okazaki,Y. and Hayashizaki,Y.

GEN Mouse ESTs
published (1999)
Contact: Chie Owa
Genome Science Laboratory
GEN

1-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
thermostabilization and thermoactivation of thermolabile enzymes by
rehealose and its application for the synthesis of full length cDNA
Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998)
transcriptional sequencing: A method for DNA sequencing using RNA
polymerase [Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998)]
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.
Location/Qualifiers
1..308
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1700058H21"
/clone_lib="Mus musculus adult C57BL/6J testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
90 a 78 c 50 g 91 t

ORIGIN

Query Match 79.2%; Score 19; DB 9; Length 308;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgactttcaactctgtct 23
|||||

Db 126 CTGACTTTCAACTCTGTCT 144
|||||

RESULT 9
BG814329/c
LOCUS

DEFINITION
daff6a08.y1 NTCHE XGC Eye1 Xenopus laevis cDNA clone IMAGE:474346
5', mRNA sequence.

ACCESSION
BG814329

VERSION
BG814329.1 GI:14185309

KEYWORDS
EST.

SOURCE
African clawed frog.

ORGANISM
Xenopus laevis

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 486)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.

AUTHORS
WashU Xenopus EST project, 1999

TITLE
WashU Xenopus EST project, 1999

JOURNAL
Other_ESTs: daff6a08.xl

COMMENT
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 444.

FEATURES
source
1..486
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:474346"
/clone_lib="NICHD XGC Eye1"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPOrr6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

BASE COUNT 174 a 87 c 173 g 52 t

ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 486;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ccctgactttcaactctgtctc 24
|||||

Db 128 CTCTGACTGTCAACTCTGTCTC 107
|||||

RESULT 10

AO329245
LOCUS

DEFINITION
nxb0044L15f CUGI Rice BAC Library Oryza sativa genomic clone
nxb0044L15f, DNA sequence.

ACCESSION
AO329245

VERSION
AO329245.1 GI:4121095

KEYWORDS
GSS.

SOURCE
Oryza sativa.

ORGANISM
Oryza sativa

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 798)
Wing,R.A. and Dean,R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)

AUTHORS
Wing,R.A. and Dean,R.A.

TITLE
A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL
Unpublished (1998)

COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence stop: 261.

FEATURES
source
1..798
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nxb0044L15f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT 143 a 222 c 97 g 336 t

ORIGIN

Query Match 78.3%; Score 18.8; DB 12; Length 798;
Best Local Similarity 90.9%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ccctgactttcaactctgtctc 24
|||||

Db 466 CCCTTACTTTCAACTCTCTCTC 487
|||||

RESULT 11
BB842527
LOCUS

DEFINITION
BB842527 RIKEN full-length enriched, 6 days neonate spleen Mus
musculus cDNA clone F43006B06 5', mRNA sequence.

BB860133

```

BASE COUNT      97 a   118 c   125 g   103 t
ORIGIN

Query Match
Best Local Similarity 76.7%; Score 18.4; DB 9; Length 443;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgactttcaactctgtctc 24
||||||| |||||||
Db 287 CTGACTTTCCACTCTGTCTC 306

RESULT 13
BE553586
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 453)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: uo83c04.x1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

MGI:1029578
Seq primer: -40RP from Gibco
High quality sequence stop: 413.
FEATURES
source
1..454
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB"
/db_xref="taxon:10090"
/clone="IMAGE:2649126"
/clone_lib="NCI-CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lohar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
```

```

BASE COUNT      104 a   116 c   119 g   115 t
ORIGIN

Query Match
Best Local Similarity 76.7%; Score 18.4; DB 9; Length 454;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgactttcaactctgtctc 24
||||||| ||||||| ||||||| |||||||
Db 214 CTGACTTTCCACTCTGTCTC 233

RESULT 15
BE866609
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 509)
Akimura, T., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Hayatsu, N., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Shibata, K., Shinagawa, A., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

MGI:1029578
Seq primer: -40RP from Gibco
High quality sequence stop: 413.
FEATURES
source
1..454
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB"
/db_xref="taxon:10090"
/clone="IMAGE:2649126"
/clone_lib="NCI-CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lohar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
```

```

BASE COUNT      104 a   119 c   120 g   110 t
ORIGIN

Query Match
Best Local Similarity 76.7%; Score 18.4; DB 10; Length 453;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgactttcaactctgtctc 24
||||||| ||||||| ||||||| |||||||
Db 247 CTGACTTTCCACTCTGTCTC 266

RESULT 14
AW211667
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 509)
Akimura, T., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Hayatsu, N., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Shibata, K., Shinagawa, A., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

MGI:1029578
Seq primer: -40RP from Gibco
High quality sequence stop: 442.
FEATURES
source
1..453
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153419"
/clone_lib="NCI-CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH
```

JOURNAL
COMMENT

2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.

FEATURES Location/Qualifiers

source

1. .509
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G4D0001E12"
/clone_lib="RIKEN full-length enriched, kidney CCL-142 RAG
cDNA"
/tissue_type="kidney"
/cell_line="CCL-142 RAG"
113 a 135 c 140 g 121 t

BASE COUNT
ORIGIN

Query Match 76.7%; Score 18.4; DB 9; Length 509;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 ctgacttcaactctgtctc 24
||||||| |||||
Db 281 CTGACTTTCACACTCTGTCTC 300

Search completed: August 24, 2002, 21:08:27
Job time: 17347 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 17:35:15 ; Search time 2613.74 Seconds
(without alignments)
240.191 Million cell updates/sec

Title: US-09-986-381-l
Perfect score: 30
Sequence: 1 gccgtttccagttgtttatctgttctact 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq-length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	30	100.0	799	9	HSP5314	
2	30	100.0	3407	9	HOMOTSP1	
3	30	100.0	3423	9	HSW0593P1	
4	30	100.0	159021	2	AC087388	
5	30	100.0	160457	2	AC008049	
6	28.4	94.7	600	6	I92482	
7	28.4	94.7	4360	9	AB018045	
8	28.4	94.7	20303	9	HSP53G	
9	28.4	94.7	20303	9	HSU94788	
10	22.2	74.0	17931	2	AC098420	
11	22	73.3	22	6	AX037125	
12	22	73.3	945	10	AY073823	
13	21.6	72.0	194529	2	AC036146	
14	21.6	72.0	200039	2	AL671880	
15	21.6	72.0	224040	2	AC023174	
16	21.2	70.7	2294	9	AK055435	
17	21.2	70.7	4428	9	AF322916	
18	21.2	70.7	4429	4	AF322915	
19	21.2	70.7	5520	9	AB046781	
20	21.2	70.7	19795	9	HSIGCMUDE	
21	21.2	70.7	83775	2	AC022466	
22	21.2	70.7	148041	2	AC087699	
23	21.2	70.7	154160	2	AC046154	
24	21.2	70.7	155760	2	AC010076	
25	21.2	70.7	167208	2	AC034190	
26	21.2	70.7	169802	9	CNS01DF2	
27	21.2	70.7	173534	9	AC007920	
28	21.2	70.7	197082	2	AC069069	
29	21.2	70.7	207841	2	AC072019	
30	21	70.0	89936	8	AP004532	
31	21	70.0	163396	9	AL161445	
32	21	70.0	176584	2	AC027812	
33	21	70.0	183099	2	AC034144	
34	21	70.0	188540	2	AC011108	
35	20.6	68.7	2633	9	HSU63041	
36	20.6	68.7	2799	9	HSNCAME	
37	20.6	68.7	2960	9	S71824	
38	20.6	68.7	3309	9	AK057509	
39	20.6	68.7	8513	6	AX207284	
40	20.6	68.7	86478	2	AC022577	
41	20.6	68.7	144135	9	AC097494	
42	20.6	68.7	165699	2	AP001802	
43	20.6	68.7	176749	9	AC096748	
44	20.6	68.7	177464	9	AP002853	
45	20.6	68.7	179598	9	AC073365	

ALIGNMENTS

RESULT	1	HSP5314	H.sapiens	intron 4	799 bp	DNA	linear	PRI 25-NOV-1996
LOCUS	DEFINITION	H.sapiens	intron 4	from p53 gene.				
ACCESSION	X92659							
VERSION	X92659.1	GI:1177472						
KEYWORDS	p53 gene.							
SOURCE	human.							
ORGANISM	human.							
REFERENCE	1	(bases 1 to 799)						
AUTHORS	Shamsher, M. and Montano, X.							
TITLE	Analysis of intron 4 of the p53 gene in human cutaneous melanoma							
JOURNAL	Gene 176 (1-2), 259-262 (1996)							
MEDLINE	97075940							
REFERENCE	2	(bases 1 to 799)						
AUTHORS	Montano, X.C.							
TITLE	Direct Submission							
JOURNAL	Submitted (29-Oct-1995) X.C. Montano, Imperial Cancer Research Fund, Viral Mediated Cell Differentiation Lab, PO Box 123, Lincoln's Inn Fields, LONDON, WC2A 3PX, UK							
FEATURES	Location/Qualifiers							


```

exon      1836..1948
/number="p53"
exon      2517..2626
/number="p53"
/number=7
exon      2970..3106
/number="p53"
/number=8
variation 2987
/number="p53"
/replace="g"
exon      3199..3272
/number="p53"
/number=9

BASE COUNT 741 a 967 c 885 g 830 t
ORIGIN

Query Match      100.0%; Score 30; DB 9; Length 3423;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgtcttcacgttgcctttatctgttcaact 30
|||||
Db 1499 GCCGCTCTCCAGTGTCTTATCTGTCACT 1528

RESULT 4
AC087388 159021 bp DNA linear HTG 29-NOV-2001
LOCUS Homo sapiens chromosome 17 clone RP11-199F11 map 17, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
AC087388 4 GI:16117576
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 159021)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-199F11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159021)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Canarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazates,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnec,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 13, 2001 this sequence version replaced gi:16041379.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11969
Center clone name: 199_F_11
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 5822: contig of 5822 bp in length
* 5823 5922: gap of 100 bp
* 5923 43817: contig of 37895 bp in length
* 43818 43917: gap of 100 bp
* 43918 119706: contig of 75789 bp in length
* 119707 119806: gap of 100 bp
* 119807 153208: contig of 33402 bp in length
* 153209 153308: gap of 100 bp
* 153309 159021: contig of 5713 bp in length.
Location/Qualifiers
1..159021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-199F11"
/clone_lib="FPC1-11 Human Male BAC"

```

FEATURES

```

BASE COUNT 40373 a 39760 c 38883 g 39366 t 639 others
ORIGIN

Query Match      100.0%; Score 30; DB 2; Length 159021;
Best Local Similarity 100.0%; Pred. No. 0.0084;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgtcttcacgttgcctttatctgttcaact 30
|||||
Db 78310 GCCGCTCTCCAGTGTCTTATCTGTCACT 78339

```

RESULT 5

```

AC008049
LOCUS Homo sapiens clone RP11-199F11, WORKING DRAFT SEQUENCE, 9 unordered
pieces
AC008049 160457 bp DNA linear HTG 17-JUL-2001
DEFINITION Homo sapiens
AC008049.37 GI:14787096
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 160457)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

```

Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherez, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleszyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y. F., Wu, J., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 160457)
Worley, K. C.

Direct Submission

Submitted (16-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 17, 2001 this sequence version replaced gi:14328991.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMGY
Center clone name: RP11-199F11
----- Summary Statistics
Sequencing vector: Plasmid; M7789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 66% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161924 bases at least Q40
Consensus quality: 166972 bases at least Q30
Consensus quality: 169776 bases at least Q20
Estimated insert size: 162538; sum-of-contrigs estimation
Quality coverage: 10.6x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 55987: contig of 55987 bp in length
* 55988 56087: gap of unknown length
* 56088 96796: contig of 40709 bp in length
* 96797 96896: gap of unknown length
* 96897 124682: contig of 27786 bp in length
* 124683 124782: gap of unknown length
* 124783 136186: contig of 11404 bp in length
* 136187 136286: gap of unknown length

* 136287 141789: contig of 5503 bp in length
* 141790 141889: gap of unknown length
* 141890 149503: contig of 7614 bp in length
* 149504 149603: gap of unknown length
* 149604 154667: contig of 5064 bp in length
* 154668 154767: gap of unknown length
* 154768 156994: contig of 2227 bp in length
* 156995 157094: gap of unknown length
* 157095 160457: contig of 3363 bp in length.

FEATURES
Location/Qualifiers
source

1..160457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-199F11"
BASE COUNT 39372 a 39284 c 40113 g 40841 t 847 others
ORIGIN

Query Match 100.0%; Score 30; DB 2; Length 160457;
Best Local Similarity 100.0%; Pred. No. 0.0084;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgtctccagttgctttatctgttccact 30
|||||

Db 72644 GCCGCTCCAGTTCGTTTATCTGTTCAC 72673

RESULT 6

192482 192482 600 bp DNA linear PAT 01-DEC-1998
LOCUS I92482 Sequence 2 from patent US 5728526.
DEFINITION I92482
ACCESSION I92482
VERSION I92482.1 GI:3936952
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 600)
AUTHORS George, A. L. Jr., Bhatnagar, S. K. and Nazarenko, I.
TITLE Method for analyzing a nucleotide sequence
JOURNAL Patent: US 5728526-A 2 17-MAR-1998;
FEATURES Location/Qualifiers
source 1..600
/organism="unknown"

BASE COUNT 102 a 170 c 168 g 160 t
ORIGIN

Query Match 94.7%; Score 28.4; DB 6; Length 600;
Best Local Similarity 96.7%; Pred. No. 0.022;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccgtctccagttgctttatctgttccact 30
|||||

Db 33 GCCGTGTTCCAGTTCGTTTATCTGTTCAC 62

RESULT 7

AB018045 4360 bp DNA linear PRI 14-APR-2000
LOCUS AB018045
DEFINITION Homo sapiens HSP70-1 gene for heat shock protein 72, spliced variant, partial cds.
ACCESSION AB018045
VERSION AB018045.1 GI:4691417
KEYWORDS HSP70-1; heat shock protein 72; HSP70-Hom; alternative splicing.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Shimizu, S., Nomura, K., Ujihara, M. and Demura, H.
TITLE An additional exon of stress-inducible heat shock protein 70 gene (HSP70-1)


```

JOURNAL      Biochem. Biophys. Res. Commun. 257 (1), 193-198 (1999)
MEDLINE      99194576
REFERENCE    2 (bases 1 to 4360)
AUTHORS      Nomura,K. and Shimizu,S.
TITLE        Direct Submission
JOURNAL      Submitted (27-SEP-1998) Kaoru Nomura, Tokyo Women's Medical
University, Department of Medicine 2; 8-1 Kawadacho, Shinjyuku-ku,
Tokyo 162-8666, Japan (E-mail:nomura7@parkcity.ne.jp.
Tel:81-3-3353-8111(ex.39223), Fax:81-3-3357-6475)
COMMENT      Sequence updated (26-Oct-1998).
FEATURES
  source     1..4360
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="6"
             /map="6p21.3"
             complement(1..196)
             /gene="HSP70-Hom"
             complement(1..196)
             /gene="HSP70-Hom"
             2323..2679
             /gene="HSP70-1"
             /note="alternative splicing"
             /number=1
             2323..4360
             /gene="HSP70-1"
             join(2632..2679,3955..>4360)
             /gene="HSP70-1"
             /note="spliced variant"
             /codon_start=1
             /product="heat shock protein 72"
             /protein_id="BAA77235.1"
             /db_xref="GI:4691418"
             /translation="MKHWPQVINDGPKVQVYKGETKAFYPEIISMWLTMKKEI
             AEAYLGVPTNAVITVPAYENDSQROATDAGVIAGNLVRLINEPTAAIAYGLDRT
             GKGRNVLIPLDGGGTFDSILLIDGIFVKATAGDTLGGEDFDNRQ"
             3443..3954
             /gene="HSP70-1"
             /note="alternative splicing
             transcription usually starts from exon 2"
             /number=2
             3955..>4360
             /gene="HSP70-1"
             /note="alternative splicing"
             /number=3
BASE COUNT   970 a 1242 c 1147 g 1001 t
ORIGIN

Query Match          94.7%; Score 28.4; DB 9; Length 4360;
Best Local Similarity 96.7%; Pred. No. 0.028;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccgtcttcacagttgctttatctgttccact 30
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 685 GCCGTGTTCCAGTGGCTTATCTGTTCCACT 714

RESULT 8
HSP53G
LOCUS        HSP53G 20303 bp DNA linear PRI 25-JUN-1997
DEFINITION  Human p53 gene for transformation related protein p53 (also called
transformation-associated protein p53, cellular tumor antigen p53,
and non-viral tumour antigen p53).
X54156
VERSION      X54156.1 GI:35213
KEYWORDS     anti-oncogene; cell cycle control; growth suppressor; heat shock
protein 70; oncogene; p53 cellular tumour antigen; p53 gene;
phosphoprotein; transforming capacity; tumor antigen.
SOURCE      Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE    1 (bases 1 to 20303)
AUTHORS      Chumakov,P.M.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-1990) Chumakov P.M., Engelhardt Inst. of
Molecular Biology, Academy of Science of the USSR, Vavilov St. 32,
117984 Moscow, USSR
REFERENCE    2 (bases 1 to 20303)
AUTHORS      Chumakov,P.M., Almazov,V.P. and Jenkins,J.R.
JOURNAL      Unpublished
REFERENCE    3 (bases 1 to 20303)
AUTHORS      Futreal,P.A., Barrett,J.C. and Wiseman,R.W.
TITLE        An Alu polymorphism intragenic to the TP53 gene
JOURNAL      Nucleic Acids Res. 19 (24), 6977 (1991)
MEDLINE      92107726
COMMENT      See also entries K03199, M14690, M14695, X01405, X02469, M22881-4,
M22887-8, M22894-8.
See also Mol. Cell. Biol. 6:1379-1385(1986);
and Mol. Cell. Biol. 7:961-963(1987).
FEATURES
  source     1..20303
             /organism="Homo sapiens"
             /strain="caucasian"
             /db_xref="taxon:9606"
             /chromosome="17"
             /map="p13"
             join(843..949,11689..11790,11906..11927,12021..12299,
             13055..13238,13320..13432,14000..14109,14452..14588,
             14681..14754,17572..17678,18599..19876)
             /gene="p53"
             /gene="p53"
             /gene="p53"
             /number=1
             843..19876
             /gene="p53"
             843..949
             /gene="p53"
             /number=1
             843..19876
             /gene="p53"
             950..11688
             /gene="p53"
             /number=1
             2581..2587
             /gene="p53"
             /note="5'-ALU flanking"
             2588..2877
             /gene="p53"
             /rpt_family="ALU"
             2890..2896
             /gene="p53"
             /note="3' ALU-flanking"
             3915..3929
             /gene="p53"
             /note="3' ALU-flanking"
             3950..4223
             /gene="p53"
             /rpt_family="ALU"
             4224..4238
             /gene="p53"
             /note="5' ALU flanking"
             4319..4327
             /gene="p53"
             /note="5'-ALU flanking"
             4328..4603
             /gene="p53"
             /rpt_family="ALU"
             4631..4639
             /gene="p53"
             /note="3' ALU-flanking"
             4786..5574
             /gene="p53"
             /note="rearranged cluster"
             /rpt_family="ALU"
             5802..5811
             /gene="p53"
             /note="5'-ALU flanking"

```


Unpublished
 REFERENCE 3 (bases 1 to 20303)
 AUTHORS Rozemuller, E.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAR-1997) Pathology, University Hospital Utrecht,
 P.O.Box 85500, Utrecht 3508GA, The Netherlands
 COMMENT This sequence describes corrections and polymorphisms in the
 genomic P53 sequence and refers to the genomic tp53 sequence with
 GenBank Accession Number X54156.
 FEATURES
 Location/Qualifiers
 1..20303
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 438
 /note="G in this sequence; GG found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="gg"
 470..473
 /note="TTTT in this sequence; TTT found in GenBank
 Accession Number X54156"
 /citation=[2]
 /replace="ttt"
 606
 /note="Polymorphism: A present or absent"
 /replace=""
 join(843..949,11689..11790,11906..11927,12021..12299,
 13055..13238,13320..13432,14000..14109,14452..14588,
 14681..14754,17572..17678,18599..19876)
 /gene="TP53"
 /product="p53"
 843..19876
 /gene="TP53"
 843..949
 /gene="TP53"
 /number=1
 11615..11616
 /gene="TP53"
 /note="GT in this sequence; CC found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="cc"
 11621
 /gene="TP53"
 /note="T in this sequence; C found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="c"
 11654
 /gene="TP53"
 /note="T in this sequence; C found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="c"
 11689..11790
 /gene="TP53"
 /number=2
 join(11717..11790,11906..11927,12021..12299,13055..13238,
 13320..13432,14000..14109,14452..14588,14681..14754,
 17572..17678,18599..19876)
 /gene="TP53"
 /codon_start=1
 /product="p53"
 /protein_id="AAC12971.1"
 /db_xref="GI:3041867"
 /translation="MEEPQSDFSVEPPLSQFTDLKLLPENNVLSPLPSQAMDDL
 LSPDIEQFDEPDEAPRMPAARVAPAPAPAPAPAPSPVSSVPSQKT
 YGSGVFLGRKTSVCTVSPALNKFQCLAKTCVQLMVDSTPPGTRVRAM
 AITYQSQRHTVVRRCPPHRCSDSLAPPOHLIRVEGNLRVYLDLDRNTFRHSVV
 PYEPFVSGDCTTHYNYMNSSCMGMRPILITLEDSSGNLLGRNFEVRVCA
 CERDRIIEENLRKKGEPHLPFGSKRALPNNTSSSPQPKKKPLDGEYFIQING
 REFEMFRELNALELKDAQAGKEPGGSGRAHSHLKSCKGQSTSRHKKLMFKTEGPD
 S"

variation
 /gene="TP53"
 /note="Polymorphism: G or C"
 /replace="c"
 11906..11927
 /gene="TP53"
 /number=3
 11948..11951
 /gene="TP53"
 /note="GGG in this sequence; GGG found in GenBank
 Accession Number X54156"
 /citation=[2]
 /replace="ggg"
 11952..11967
 /gene="TP53"
 /note="Polymorphism: repeat polymorphism of the
 (ACCTGAGGCTGGG)n motif (n=1,2)"
 11964..11969
 /gene="TP53"
 /note="GGGG in this sequence; GGGGG found in GenBank
 Accession Number X54156"
 /citation=[2]
 /replace="ggggggg"
 12021..12299
 /gene="TP53"
 /number=4
 12032
 /gene="TP53"
 /note="Polymorphism: G or A"
 /replace="a"
 12139
 /gene="TP53"
 /note="Polymorphism: G or C"
 /replace="c"
 13055..13238
 /gene="TP53"
 /number=5
 13320..13432
 /gene="TP53"
 /number=6
 13399
 /gene="TP53"
 /note="Polymorphism: A or G"
 /replace="g"
 14000..14109
 /gene="TP53"
 /number=7
 14452..14588
 /gene="TP53"
 /number=8
 14681..14754
 /gene="TP53"
 /number=9
 17536
 /gene="TP53"
 /note="T in this sequence; A found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="a"
 17572..17678
 /gene="TP53"
 /number=10
 17699
 /gene="TP53"
 /note="C in this sequence; A found in GenBank Accession
 Number X54156"
 /citation=[2]
 /replace="a"
 18599..19876
 /gene="TP53"
 /number=11
 18717
 /gene="TP53"
 conflict

/note="A in this sequence; T found in GenBank Accession Number X54156"
/citation=[2]
/replace="t"
BASE COUNT 5245 a 4967 c 5107 g 4984 t
ORIGIN

Query Match 94.7%; Score 28.4; DB 9; Length 20303;
Best Local Similarity 96.7%; Pred. No. 0.035, Mismatches 0; Indels 0; Gaps 0;
Matches 29; Conservative 0

QY 1 gcgcgtccagttgcttattctgttcaact 30
||||| ||||||| ||||||| ||||||| |||||||
Db 12983 GCCGTGTTCCAGTGTCTTATCTGTTCACT 13012

RESULT 10
AC098420
LOCUS
DEFINITION Rattus norvegicus clone CH230-2E23, *** SEQUENCING IN PROGRESS ***, linear HTG 23-OCT-2001
AC098420
VERSION AC098420.1 GI:16328187
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 17931)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.J., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,I., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,D., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Slisov,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs.R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 17931)
TITLE JOURNAL
REFERENCE

AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TURK
Center clone name: CH230-2E23
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 12471 bases at least Q40
Consensus quality: 14280 bases at least Q30
Consensus quality: 15385 bases at least Q20
Estimated insert size: 41807; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft_data.html).
* NOTE: this is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1
* 3027 3026: contig of 3026 bp in length
* 3127 3126: gap of unknown length
* 5170 5169: contig of 2043 bp in length
* 5270 5269: gap of unknown length
* 9601 9600: contig of 4331 bp in length
* 9701 9700: gap of unknown length
* 12729 12728: contig of 3028 bp in length
* 12829 12828: gap of unknown length
* 15441 15441: contig of 2613 bp in length
* 15442 15441: gap of unknown length
* 15542 17931: contig of 2390 bp in length.
* Location/Qualifiers
* 1..17931
* /organism="Rattus norvegicus"
* /db_xref="taxon:10116"
* /clone="CH230-2E23"
BASE COUNT 5197 a 3268 c 3511 g 5450 t 505 others
ORIGIN

Query Match 74.0%; Score 22.2; DB 2; Length 17931;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gtcttcagttgcttattctgttcaact 30
||||| ||||||| ||||||| ||||||| |||||||
Db 11974 GTTTCAGTGTCTTATCTGTTCACT 12000

RESULT 11
AX037125
LOCUS AX037125
DEFINITION Sequence 37 from Patent WO0056923.
ACCESSION AX037125
VERSION AX037125.1 GI:11226550
KEYWORDS
SOURCE Synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 22)
AUTHORS Sibson,R.

TITLE Genetic analysis
JOURNAL Patent: WO 0056923-A 37 28-SEP-2000;
SIBSON ROSS (GB) : CLATTERBRIDGE CANCER RES TRUST (GB)
FEATURES Location/Qualifiers
source i. .22
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PCR primer"
BASE COUNT 3 a 5 c 3 g 11 t
ORIGIN

Query Match 73.3%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 ttccagttgctttatctgttca 28
|||||
Db 1 TTCCAGTTGCTTTATCTGTTCA 22

RESULT 12
AY073823
LOCUS 945 bp DNA linear ROD 04-FEB-2002
DEFINITION Mus musculus olfactory receptor MOR267-16 gene, complete cds.
ACCESSION AY073823
VERSION AY073823.1 GI:18480943
KEYWORDS house mouse.
SOURCE Mus musculus

REFERENCE 1
AUTHORS Zhang,X. and Firestein,S.
TITLE The olfactory receptor gene superfamily of the mouse
JOURNAL Nat. Neurosci. 5 (2), 124-133 (2002)
PUBMED 11802173
REFERENCE 2 (bases 1 to 945)
AUTHORS Adams,M.
TITLE Direct Submission

JOURNAL Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1. .945
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="8"

mRNA <1. .>945
CDS 1. .945
/product="olfactory receptor MOR267-16"

/evidence=not_experimental
/product="olfactory receptor MOR267-16"
/protein_id="AA61486.1"
/db_xref="GI:1848094"
/translation="MECVNDTVVREFVFLGSSLAELQLLLFAVLSLYLTSTNAV
IVSTVLDRALHPMPFLSVLSCSTCYTFVIVPKMLVDLLARKKSISFLGCAIQMF
TFPLGSSHFLAAMGYDRYVAICHPRLTYVLMGRVCGVLAACVCGTVAQVIT
SQVRLPFRSSNLHFFCDISPLVQLASHPHSTQITIFLLCALVIVIPLLILVSY
IHISAILQFPSTLGRYKAFSTCASHDIVIVVHYGCASFYILRPKSYSSSQDALISV
SYTILTPFPNFIYSLRNKDFKSALHRVIGRTVTLRQH"
BASE COUNT 178 a 286 c 186 g 295 t
ORIGIN

Query Match 73.3%; Score 22; DB 10; Length 945;
Best Local Similarity 83.3%; Pred. No. 21;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gccgtctccagttgctttatctgttca 30
|||||
Db 85 GCCGCTTCCGTGCCCTCTATTGTTCACT 114

RESULT 13
AC036146/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC036146 194529 bp DNA linear HTG 11-JUN-2001
Mus musculus chromosome 5 clone RP23-294A15 map 5, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC036146
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194529)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus chromosome 5, clone RP23-294A15
Unpublished

2 (bases 1 to 194529)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 11, 2001 this sequence version replaced gi:7960432.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8871

Center clone name: 294_A15

----- Summary Statistics

Sequencing vector: M13; M77815; 45% of reads
Sequencing vector: Plasmid; n/a; 54% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 193311 bases at least Q40
Consensus quality: 193785 bases at least Q30
Consensus quality: 193926 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 194029; sum-of-contigs
Quality coverage: 9.7 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.

TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Unpublished

2 (bases 1 to 224040)

Green, E.D.

Direct Submission

Submitted (09-FEB-2000) NIH Intramural Sequencing Center, 8717

Grovenmont Circle, Gaithersburg, MD 20877, USA

On Oct 27, 2001 this sequence version replaced gi:6957694.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@hgri.nih.gov

----- Project Information

Center project name: yz

Center clone name: 025N12

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 219345 bases at least Q40

Consensus quality: 220414 bases at least Q30

Consensus quality: 220834 bases at least Q20

Insert size: 205000; agarose-fp

Insert size: 228000; pulse-field-gel

Insert size: 223440; sum-of-contigs

Quality coverage: 9.50x in Q20 bases; agarose-fp

Quality coverage: 8.54x in Q20 bases; pulse-field-gel

Quality coverage: 8.72x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 6890: contig of 6890 bp in length
* 6891
* 6990: gap of unknown length
* 6991
* 19055: contig of 12065 bp in length
* 19056
* 19155: gap of unknown length
* 19156
* 52333: contig of 33178 bp in length
* 52334
* 52433: gap of unknown length
* 52434
* 77493: contig of 25060 bp in length
* 77494
* 77593: gap of unknown length
* 77594
* 101540: contig of 23947 bp in length
* 101541
* 101540: gap of unknown length
* 101541
* 158652: contig of 57012 bp in length
* 158653
* 158752: gap of unknown length
* 158753
* 224040: contig of 65288 bp in length.

FEATURES

Source

1. .224040
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="6"
/clone="RP23-25N12"
/clone_lib="RPC1 mouse BAC library 23"

misc_feature

1. .6890
/note="assembly_fragment"
clone_end:T7
vector_side:left"

misc_feature

6991..19055
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

misc_feature

19156..52333

misc_feature /note="assembly_fragment"
52434..77493
misc_feature /note="assembly_fragment"
77594..101540
misc_feature /note="assembly_fragment"
101641..158652
misc_feature /note="assembly_fragment"
158753..224040
misc_feature /note="assembly_fragment"

BASE COUNT 65716 a 46008 c 46116 g 65377 t 823 others
ORIGIN

Query Match 72.0%; Score 21.6; DB 2; Length 224040;
Best Local Similarity 85.7%; Pred No. 67;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctccagttgctttatctgttcaact 30
| ||||| || ||||| ||||| |||||
Db 116337 CTCTCCATTTCTTTATTGTTCACT 116310

Search completed: August 24, 2002, 21:52:24
Job time: 15429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 24, 2002, 18:52:30 ; Search time 905.65 Seconds
(without alignments)
56.873 Million cell updates/sec

Title: US-09-986-381-l
Perfect score: 30
Sequence: 1 gccgtttccagtggtttatctgttcaact 30

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_032802.*			
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*						
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*						
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*						
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*						
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*						
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*						
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*						
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*						
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*						
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*						
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*						
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*						
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*						
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*						
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*						
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*						
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*						
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*						
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*						
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*						
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*						
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*						
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*						
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28.4	94.7	600	AAT59674	Human p53 gene fra
2	22	73.3	22	AAC63324	Human p53 exon 5 p
C 3	20.6	68.7	4710	AAH98595	Human EST-derived
C 4	20.6	68.7	4710	AAH98610	Human EST-derived
C 5	20.6	68.7	8513	AAD14203	MESVR/EGFP/IRESNCA
C 6	19.6	65.3	1980	AAK40197	Sequence of C3Vs g
C 7	19.6	65.3	4643	ABK09729	Human ovarian tumo
8	19.4	64.7	2213	AAZ60683	DNA encoding a mur
9	19	63.3	458	AAK55605	Human immune/haema

10	19	63.3	938	22	AAK66406	Human immune/haema
11	19	63.3	938	22	AAK69370	Human immune/haema
12	19	63.3	1275	17	AAT13704	ACNPV ORF 114, res
13	19	63.3	133894	17	AAT13635	ACNPV genomic DNA
C 14	18.8	62.7	1109	21	AAZ27931	Corn phosphatidylg
C 15	18.8	62.7	4965	22	AAH98348	Human EST-derived
C 16	18.8	62.7	4968	22	AAT60204	Human polynucleoti
C 17	18.6	62.0	366	20	AAV89531	EST clone CP314.
C 18	18.6	62.0	521	21	AAC04378	Human secreted pro
C 19	18.6	62.0	637	23	ABL05333	Drosophila melanog
C 20	18.6	62.0	1140	21	AAK66003	E. coli proliferat
C 21	18.6	62.0	2574	20	AAK07111	Staphylococcus aur
C 22	18.6	62.0	2574	23	AAK52524	E. coli DNA for ce
C 23	18.6	62.0	2696	23	ABL05332	Drosophila melanog
C 24	18.6	62.0	2917	23	ABL05604	Drosophila melanog
C 25	18.6	62.0	3016	23	ABL05328	Drosophila melanog
C 26	18.6	62.0	3409	23	ABL05530	Drosophila melanog
C 27	18.6	62.0	3503	22	AAI66060	Escherichia coli C
C 28	18.4	61.3	596	22	AAK03903	Human secreted pro
C 29	18.4	61.3	2225	21	AAZ52467	HTRM clone 1880692
C 30	18.4	61.3	2231	23	ABL07347	Drosophila melanog
C 31	18.4	61.3	2459	21	AAK75871	Human ORFX ORF1426
C 32	18.4	61.3	3203	22	AAH17998	Human cDNA sequenc
C 33	18.4	61.3	3237	22	AAH15741	Human cDNA sequenc
C 34	18.4	61.3	3314	22	AAH14681	Human vesicle traf
C 35	18.4	61.3	3387	22	AAK09389	Human immune/haema
C 36	18.4	61.3	4472	23	ABL07346	Drosophila melanog
C 37	18.4	61.3	18660	22	AAK79108	Human digestive sy
C 38	18.4	61.3	18671	22	AAK90763	Human immune/haema
C 39	18.4	61.3	31584	22	AAK81054	Human immune/haema
C 40	18.4	61.3	42432	21	AAK55187	Cenarchaeum symbio
C 41	18.2	60.7	644	23	AAK64918	DNA encoding novel
C 42	18.2	60.7	1402	22	AAK03926	Corn NADH oxidase-
C 43	18.2	60.7	1701	23	AAK73392	DNA encoding novel
C 44	18.2	60.7	1701	23	AAK593941	DNA encoding novel
C 45	18.2	60.7	3047	21	AAZ52355	NS5Q gene-7 associ

ALIGNMENTS

RESULT	1
AAT59674	
ID	AAT59674 standard; DNA; 600 BP.
XX	
AC	AAT59674;
XX	
DT	07-OCT-1997 (first entry)
XX	
DE	Human p53 gene fragment.
XX	
KW	Target nucleotide sequence; point mutation; analysis;
KW	criminal investigation; parental determination; ss.
OS	Homo sapiens.
XX	
EH	Key
FT	Location/Qualifiers
FT	105..288
FT	/*tag= a
FT	/number= 1
FT	/note= "Exon in which possible mutation sites occur"
FT	misc_difference 163
FT	/*tag= b
FT	/note= "Wild-type T could be point mutated to G"
FT	misc_difference 267
FT	/*tag= c
FT	/note= "Wild-type G could be point mutated to C"
XX	
PN	WO9641001-A1.
XX	
PD	19-DEC-1996.
XX	
PF	05-JUN-1996; 96WO-US08759.

```
XX 07-JUN-1995; 95US-0472239.
XX (ONCO-) ONCOR INC.
XX Bhatnagar SK, George AL, Nazarenko I;
XX WPI; 1997-099941/09.
XX Analysis of target nucleotide sequence which exits in a first state
PT or different second state - useful for determination of point
PT mutation(s)
XX Disclosure; Page 46; 78pp; English.
XX A novel method has been produced for analysing a target nucleotide
CC sequence. The target sequence having a first segment, a second segment
CC and a third segment between the first and second, which is formed of at
CC least one, but less than four different, nucleotide(s) where the third
CC segment has a nucleotide or nucleotide sequence in a first state or
CC different second state. The present sequence is a partial human p53
CC gene sequence which is used as an example of a target nucleotide
CC sequence. The method can be used to analyse the entire sequence
CC (portion) of a known gene and to analyse infectious disease or to
CC determine if a sample is from a particular source, such as for criminal
CC investigations or parental determination. Especially the method is used
CC to determine the existence or absence of mutations consisting of one or
CC a few nucleotides.
XX Sequence 600 BP; 102 A; 170 C; 168 G; 160 T; 0 other;
SQ

Query Match 94.7%; Score 28.4; DB 18; Length 600;
Best Local Similarity 96.7%; Pred. No. 0.018;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccgttccagtgctttatctgttcaact 30
Db 33 gccgttccagtgctttatctgttcaact 62

RESULT 2
AAC63324
ID AAC63324 standard; DNA; 22 BP.
AC AAC63324;
XX 06-FEB-2001 (first entry)
XX Human p53 exon 5 PCR primer #1.
XX Human; mutation detection; polynucleotide sequence comparison;
KW hybridisation assay; modification library; PCR primer; ss.
XX Homo sapiens.
OS Synthetic.
XX GB2348284-A.
XX 27-SEP-2000.
XX 23-NOV-1999; 99GB-0027520.
XX 24-MAR-1999; 99GB-0006833.
XX (CLAT-) CLATTERBRIDGE CANCER RES TRUST.
XX Sibson R;
XX WPI; 2000-595847/57.
XX Comparing polynucleotide (PN) sequences for determining the nature
PT and/or position of an unknown difference between PNs, comprises
```

```
PT hybridization assays using modification libraries of sub-populations of
PT PNS
XX Example 3; Page 67; 111pp; English.
XX The present sequence was used in a new method for comparing
CC polynucleotide sequences. The method uses hybridisation assays and
CC modification libraries of two sub-populations of polynucleotides.
CC It can be used for determining the nature and/or position of an unknown
CC difference between polynucleotides. The method provides a high throughput
CC method of sequence analysis which has the sensitivity to detect both the
CC positions and identities of sequence changes right down to single base
CC resolution and is able to discriminate variants when they exist as only
CC a small proportion of the total sequences in a sample. The methods
CC employ fewer steps than prior known methods. The amount of sample
CC reaction products does not require specialised systems. Multiple sequence
CC positions can be analysed at the same time in a single solution.
XX Sequence 22 BP; 3 A; 5 C; 3 G; 11 T; 0 other;
SQ

Query Match 73.3%; Score 22; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ttccagtgctttatctgttca 28
Db 1 ttccagtgctttatctgttca 22

RESULT 3
AAH98595/c
ID AAH98595 standard; cDNA; 4710 BP.
XX AAH98595;
AC AAH98595;
XX 12-OCT-2001 (first entry)
XX Human EST-derived coding sequence SEQ ID NO: 452.
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX Homo sapiens.
XX WO200154477-A2.
XX 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US02687.
XX 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
XX P-PSDB; AAM23936.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX Claim 1; Page 500-501; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
```

CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 4710 BP; 1179 A; 1258 C; 1167 G; 1106 T; 0 other;

Query Match 68.7%; Score 20.6; DB 22; Length 4710;
Best Local Similarity 85.2%; Pred. No. 44;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtcttcagtgctttatctgttcac 29
DB 923 CGTCTTCTCTGCTCTATCTGTTCCC 897

RESULT 4
AAH98610/c
ID AAH98610 standard; cDNA; 4710 BP.
XX
AC AAH98610;

DT 12-OCT-2001 (first entry)

XX Human EST-derived coding sequence SEQ ID NO: 467.

DE Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

OS WO200154477-A2.

PN 02-AUG-2001.

PD 25-JAN-2001; 2001WO-US02687.

XX 25-JAN-2000; 2000US-0491404.

XX 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

DR P-PSDB; AAM23951.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

XX Claim 1; Page 509-510; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

XX Sequence 4710 BP; 1179 A; 1258 C; 1167 G; 1106 T; 0 other;

Query Match 68.7%; Score 20.6; DB 22; Length 4710;
Best Local Similarity 85.2%; Pred. No. 44;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtcttcagtgctttatctgttcac 29
DB 923 CGTCTTCTCTGCTCTATCTGTTCCC 897

RESULT 5
AAD14203/c
ID AAD14203 standard; DNA; 8513 BP.
XX
AC AAD14203;

DT 06-NOV-2001 (first entry)

XX MESVR/EGFP/IRESNCAmPro(ori) vector.

DE Transcriptional regulatory element; translational regulatory element;
XX gene expression; diagnostic application; therapeutic application;
KW MESVR; retrovirus vector; enhanced green fluorescent protein; EGFP; IRES;
KW internal ribosome entry sequence; N-CAM; neural cell adhesion molecule;
KW ori; origin of replication; SV40; human; mouse; ds.

OS Chimeric - Murine embryonic stem cell virus.

OS Chimeric - Rous sarcoma virus.

OS Chimeric - Mastadenovirus.

OS Chimeric - Mus sp.

OS Chimeric - Rhesus macaque polyoma virus.

OS Chimeric - Homo sapiens.

XX Chimeric - Synthetic.

PN WO200155371-A1.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-US02733.

XX 28-JAN-2000; 2000US-0178816.

PR 02-MAR-2000; 2000US-0186496.

PR 30-MAY-2000; 2000US-0207804.

PR 07-SEP-2000; 2000US-0230852.

PR 07-SEP-2000; 2000US-0230956.

PR 12-JAN-2001; 2001US-0261312.

XX (SCRI) SCRIPPS RES INST.

XX Mauro VP, Edelman GM, Chappell GM, Jones FS, Owens G, Meech R;

XX WPI; 2001-483242/52.

XX Identifying oligonucleotides with transcriptional/translational
PT regulatory activity in eukaryotic cells by integrating an
PT oligonucleotide into cell genome and detecting a change in expression
PT of expressible polynucleotides -

XX Claim 95; Page 131-135; 172pp; English.

XX The present invention relates to a method for the identification of an
CC oligonucleotide with transcriptional/translational regulatory activity in
CC a eukaryotic cell. The method involves integrating an oligonucleotide
CC into the genome of a eukaryotic cell so that it is linked to an
CC expressible polynucleotide or contacting the eukaryotic cell with library
CC of vectors, obtained by cloning a library of oligonucleotide into
CC multiple copies of expression vectors comprising polynucleotide, and
CC detecting changes in expression of the polynucleotide. The method is
CC useful for identifying an oligonucleotide having transcriptional or
CC translational activity in a eukaryotic cell. It is also useful for
CC identifying synthetic transcriptional or translational regulatory
CC elements. The transcriptional or translational regulatory elements are
CC useful in a variety of gene expression configurations for regulating
CC control of expression, and in expression vectors for controlling gene

CC expressions in diagnostic and therapeutic applications. The present
 CC sequence is MESVR/EGFP/IRESCAMP(ori) vector based on
 CC murine embryonic stem cell virus (MESV) retrovirus. The vector comprises
 CC nucleotide sequences encoding enhanced green fluorescent protein
 CC (EGFP) and human neural cell adhesion molecule (N-CAM) linked by an
 CC internal ribosome entry sequence (IRES). The upstream long terminal
 CC repeat (LTR) region of the retrovirus was modified to contain
 CC Rous sarcoma virus (RSV) enhancer elements. The downstream LTR U3 region
 CC was modified to contain a cassette containing a polylinker for the
 CC insertion of random oligonucleotides, adenovirus major late promoter,
 CC the initiator sequence (Inr) from mouse terminal deoxynucleotidyl
 CC transferase gene and a complete R region. The vector also contains a
 CC simian virus 40 (SV40) origin of replication (ori) and is useful
 CC for identifying synthetic transcriptional regulatory elements.
 XX
 SQ Sequence 8513 BP; 2067 A; 2353 C; 2224 G; 1869 T; 0 other;

Query Match 68.7%; Score 20.6; DB 22; Length 8513;
 Best Local Similarity 85.2%; Pred. No. 48;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctccagttgcttattctgttcc 29
 ||||| ||||| ||||| ||||| |||||
 Db 3694 CGTCTCTCTGCTATCTATCTGTTCC 3668

RESULT 6
 AAX40197/c
 ID AAX40197 standard; DNA; 1980 BP.
 XX
 AC AAX40197;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Sequence of C3Vs gene.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX
 DR WPI; 1999-132448/11.
 DR P-PSDB; AAY06996.
 XX
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Claim 67; Page 776-777; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 1980 BP; 683 A; 351 C; 511 G; 435 T; 0 other;

Query Match 65.3%; Score 19.6; DB 20; Length 1980;
 Best Local Similarity 84.6%; Pred. No. 1e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctccagttgcttattctgttca 28
 ||||| ||||| ||||| ||||| |||||
 Db 1408 CGTCTCTTAACGTGCTTATCTGCTCA 1383

RESULT 7
 ABK09729/c
 ID ABK09729 standard; cDNA; 4643 BP.
 XX
 AC ABK09729;
 XX
 DT 14-MAR-2002 (first entry)
 XX
 DE Human ovarian tumour protein encoding cDNA #262.
 XX
 KW Human; ovarian tumour protein; cancer; cytostatic; immunostimulant; ss;
 KW gene therapy; CD4+ T cell; CD8+ T cell; PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN W0200190154-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 23-MAY-2001; 2001WO-US16895.
 XX
 PR 24-MAY-2000; 2000US-207107P.
 PR 13-JUN-2000; 2000US-211457P.
 PR 21-JUN-2000; 2000US-213673P.
 PR 03-AUG-2000; 2000US-223288P.
 PR 01-MAR-2001; 2001US-272790P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Mitcham JL, Harlocker SL, Dillon DC, Secrist H, Lodes MJ;
 PI Algate PA, Fling SP, Mannion J, Benson DR, Carter D;
 XX
 DR WPI; 2002-097641/13.
 XX
 XX New isolated polynucleotide encoding polypeptide comprising portion of
 PT ovarian tumour protein, useful for detection, diagnosis and therapy of
 PT human ovarian cancer -
 XX
 PS Claim 1; Page 210-212; 285pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding a
 CC polypeptide comprising a portion of an ovarian tumour protein. The
 CC sequences of the invention are useful for stimulating an immune response
 CC and for treating ovarian cancer in a patient. An antigen presenting cell
 CC that expresses the sequences is useful for treating ovarian cancer by
 CC incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells
 CC can then be proliferated and administered to the patient to inhibit the
 CC development of cancer. The DNA sequences are useful as probes or primers

CC for nucleic acid hybridisation, to direct expression of a polypeptide in
CC appropriate host cells. Detecting the presence of a cancer in a patient
CC involves obtaining a biological sample from the patient, contacting the
CC biological sample with an agent that binds to the protein, detecting the
CC amount of protein that binds to the agent, comparing the amount of
CC protein to a predetermined cut-off value and determining the presence of
CC cancer. Sequences ABK09464-ABK09802 represent PCR primers and cDNA
CC molecules encoding ovarian tumour proteins of the invention.
XX
SQ Sequence 4643 BP; 1645 A; 872 C; 1185 G; 941 T; 0 other;

Query Match 65.3%; Score 19.6; DB 24; Length 4643;
Best Local Similarity 84.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cgtcttcagtgcttctatctgttca 28
||||| | ||||| ||||| |||
Db 1735 CGTCTTCTAACTGCTTTATCTGCTCA 1710

RESULT 8
AAZ60683
ID AAZ60683 standard; DNA; 2213 BP.

AC AAZ60683;

DT 16-MAY-2000 (first entry)

DE DNA encoding a murine caspase-like polypeptide.

XX Mouse; caspase-like polypeptide; human; caspase; apoptosis;
KW skin disease; keratinisation; wound healing; ss.

OS Mus musculus.

FH Key Location/Qualifiers
FT CDS 81..854
/*tag= a
/product= "caspase-like polypeptide"

PN WO200004169-A1.

PD 27-JAN-2000.

PF 12-JUL-1999; 99WO-EP04939.

PR 17-JUL-1998; 98EP-0202422.

PA (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PI Van De Craen M, Declercq W, Vandenabeele P, Fiers W;

XX WPI; 2000-182433/16.

DR P-PSDB; AAY68864.

XX New murine and human caspase homologues useful for treating skin
PT related disorders -

XX Claim 4; Page 48-50; 68pp; English.

PS The present sequence encodes a murine caspase-like polypeptide. The
CC specification also describes a human caspase-like polypeptide.
CC Caspases are cysteinyl aspartate-specific proteinases which play a
CC central role in apoptosis. The polypeptides of the invention are related
CC to human and murine caspase-2 and human caspase-9, and possess all of
CC the typical amino acids involved in catalysis, including the QACRG box,
CC and contain no or only a very short prodomain. mRNA expression of the
CC homologues of the invention is predominant in the skin. The caspase-like
CC polypeptides are useful for treating human or animal diseases, such
CC as skin diseases. They are also useful for screening for compounds that
CC modulate its activity, i.e. agonists, antagonists, and inhibitors. The
CC caspase-like polypeptides and polynucleotides are useful for modulating

CC keratinisation, for diagnosing and treating inappropriate wound
CC healing.

XX Sequence 2213 BP; 641 A; 573 C; 458 G; 541 T; 0 other;

SQ Query Match 64.7%; Score 19.4; DB 21; Length 2213;
Best Local Similarity 79.3%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcgcgttcagtgcttctatctgttca 29
||||| | ||||| ||||| |||
Db 872 gcttctcagtgcttctctctgttgac 900

RESULT 9
AAK55605
ID AAK55605 standard; cDNA; 458 BP.

XX AAK55605;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:665.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.

OS Homo sapiens.

PN WO200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232198.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-483426/52.
DR P-PSDB; AAM82824.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 1; SEQ ID NO 665; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 458 BP; 134 A; 95 C; 100 G; 126 T; 3 other;

Query Match 63.3%; Score 19; DB 22; Length 458;
Best Local Similarity 81.5%; Pred. NO. 1.4e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 gtcttcacagttgctttatctgttctact 30
Db | | | | | | | | | | | | | | | | | | | |
4 gacttcacagttcttcttcttcttctact 30
RESULT 10
AAK66406
ID AAK66406 standard; DNA; 938 BP.

XX	AAK66406;	PR	14-SEP-2000;	2000US-0232400.
AC		PR	14-SEP-2000;	2000US-0232401.
XX		PR	14-SEP-2000;	2000US-0233063.
DT	06-NOV-2001 (first entry)	PR	14-SEP-2000;	2000US-0233064.
XX		PR	14-SEP-2000;	2000US-0233065.
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21218.	PR	21-SEP-2000;	2000US-0234223.
XX		PR	21-SEP-2000;	2000US-0234274.
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	PR	25-SEP-2000;	2000US-0234997.
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	PR	25-SEP-2000;	2000US-0234998.
XX		PR	26-SEP-2000;	2000US-0235484.
XX	Homo sapiens.	PR	27-SEP-2000;	2000US-0235834.
XX		PR	27-SEP-2000;	2000US-0235836.
PN	WO200157182-A2.	PR	29-SEP-2000;	2000US-0236327.
XX		PR	29-SEP-2000;	2000US-0236367.
XX		PR	29-SEP-2000;	2000US-0236368.
PD		PR	29-SEP-2000;	2000US-0236369.
XX		PR	29-SEP-2000;	2000US-0236370.
PF	17-JAN-2001; 2001WO-US01354.	PR	02-OCT-2000;	2000US-0236802.
XX		PR	02-OCT-2000;	2000US-0237037.
XX	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000;	2000US-0237038.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000;	2000US-0237039.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000;	2000US-0237040.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000;	2000US-0239935.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000;	2000US-0239937.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000;	2000US-0240960.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000;	2000US-0241221.
PR	19-MAY-2000; 2000US-0205151.	PR	20-OCT-2000;	2000US-0241785.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000;	2000US-0241786.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000;	2000US-0241787.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000;	2000US-0241809.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000; 2000US-0217487.	PR	01-NOV-2000;	2000US-0244617.
PR	11-JUL-2000; 2000US-0217496.	PR	08-NOV-2000;	2000US-0246474.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000;	2000US-0246476.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000;	2000US-0246611.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000;	2000US-0246613.
PR	18-AUG-2000; 2000US-0226279.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000; 2000US-0226686.	PR	17-NOV-2000;	2000US-0249209.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000;	2000US-0249210.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000;	2000US-0249211.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000;	2000US-0249215.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000;	2000US-0249217.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000;	2000US-0249244.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000; 2000US-0232080.	PR	01-DEC-2000;	2000US-0250160.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000; 2000US-0231968.	PR	03-DEC-2000;	2000US-0251030.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000; 2000US-0232399.			

```
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 21218; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 938 BP; 259 A; 201 C; 201 G; 277 T; 0 other;

Query Match          63.3%; Score 19; DB 22; Length 938;
Best Local Similarity 81.5%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 gttctccagttgctttatctgttcaact 30
   | ||||| |||| || |||||
Db 55 gacttccagtttcttctcattcaact 81

RESULT 11
AAK69370
ID AAK69370 standard; DNA; 938 BP.
XX
XX AAK69370;
AC
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24182.
XX
XX Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
```


PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX

PS Disclosure; SEQ ID NO 24182; 307lpp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients' own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 938 BP; 259 A; 201 C; 201 G; 277 T; 0 other;

Query Match 63.3%; Score 19; DB 22; Length 938;
Best Local Similarity 81.5%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gtcttcagttgctttatctgttctact 30
| ||||| ||||| || |||||
Db 55 gacttcagtttcttctcattctact 81

RESULT 12
AAT13704
ID AAT13704 standard; DNA; 1275 BP.
XX
AC AAT13704;
XX
DT 05-SEP-1996 (first entry)
XX
DE AcNPV ORF 114, residues 99160-97886.
XX
KW Autographa californica nuclear polyhedrosis virus clone 6;
KW disruption; non-essential gene; heterologous protein production;
KW expression vector; baculovirus; ss.
XX
OS Autographa californica nuclear polyhedrosis virus clone 6.
XX
FH Key Location/Qualifiers
CDS 1..1275
FT /*tag= a
FT /number= ORF 114
FT /note= "corresponds to AcNPV nucleotides
FT complement (99160-97886)";
XX
XX WO9601320-A2.
XX
XX 18-JAN-1996.
XX
XX 30-JUN-1995; 95WO-IB00578.
XX
XX 04-JUL-1994; 94GB-0013420.
XX
XX (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
XX
XX Ayres M, Bishop D, Possee R;
XX
XX WPI; 1996-087670/09.
DR GENBANK; L22858.
XX
XX Autographa californica nuclear polyhedrosis virus complete genome
PT sequence - useful in the prodn. of vectors for enhanced
PT heterologous protein expression, such as interleukin(s),
PT

PT Interferon(s) and neurotoxin(s)
XX PS Claim 1; Page 90-186; 122pp; English.
XX
CC AAT13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46.
CC 50-60, 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130,
CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified in
CC the Autographa californica nuclear polyhedrosis virus (AcNPV) clone 6.
CC Each gene is numbered according to its position in the virus genome
CC beginning at the left end of the linear map, and irrespective of its
CC orientation. The direction of transcription is relative to that of the
CC polyhedrin gene. Of the 154 ORFs identified it was found that some of the
CC ORFs (ORFs 27, 30, 32, 71, 86, 123, 126 and 127) are dispensable for
CC virus replication in cell culture or insect larvae. These genes can be
CC deleted from the genome to: (a) provide additional sites for inserting
CC single or multiple copies of foreign genes; and (b) to reduce the size of
CC the virus complementary strand relative to the polyhedrin gene. The
CC present sequence is designated ORF 114, and is on the complementary
CC strand relative to the polyhedrin gene.
XX
SQ Sequence 1275 BP; 335 A; 225 C; 259 G; 456 T; 0 other;

Query Match 63.3%; Score 19; DB 17; Length 1275;
Best Local Similarity 81.5%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gccgtcttcacagtgctttatctgttc 27
|||||
DB 925 gccgtcttcacagtgctgtcagttc 951
|||||

RESULT 13
AAT13635
ID AAT13635 standard; DNA; 133894 BP.
XX
AC AAT13635;
XX
DT 03-SEP-1996 (first entry)
XX
DE AcNPV genomic DNA clone 6.
XX
KW Autographa californica nuclear polyhedrosis virus clone 6;
KW disruption; non-essential gene; heterologous protein production;
KW expression vector; baculovirus; ss.
XX
OS Autographa californica nuclear polyhedrosis virus clone 6.
XX
PN WC9601320-A2.
XX
PD 18-JAN-1996.
XX
PF 30-JUN-1995; 95WO-IB00578.
XX
PR 04-JUL-1994; 94GB-0013420.
XX
PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
XX
PI Ayres M, Bishop D, Possee R;
XX
DR WPI: 1996-087670/09.
DR GENBANK: L22858.
XX
XX Autographa californica nuclear polyhedrosis virus complete genome
PT sequence - useful in the prodn. of vectors for enhanced
PT heterologous protein expression, such as interleukin(s),
PT interferon(s) and neurotoxin(s)
XX
PS Disclosure; Page 90-186; 122pp; English.
XX
CC The complete nucleotide sequence of the genome of clone 6 of the
CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC has been determined. The sequence is taken from the Genbank record

CC L22858. The patent specification claims a polynucleotide selected from
CC open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC the patentees. See T13636-731. Expression vectors contg. the complete
CC genomic sequence of AcNPV, with the exception that at least one non-
CC essential ORF is disrupted or replaced are useful for the synthesis of
CC heterologous proteins.
XX
SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 other;

Query Match 63.3%; Score 19; DB 17; Length 133894;
Best Local Similarity 81.5%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gccgtcttcacagtgctttatctgttc 27
|||||
DB 98810 gccgtcttcacagtgctgtcagttc 98836
|||||

RESULT 14
AAA27931/C
ID AAA27931 standard; cDNA; 1109 BP.
XX
AC AAA27931;
XX
DT 12-SEP-2000 (first entry)
XX
DE Corn phosphatidylglycerophosphate synthase cDNA.
XX
KW Corn; maize; phosphatidylglycerophosphate synthase; phospholipid;
KW transgenic plant; herbicide; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 73..699
FT /*tag= a
XX
PN W0200036117-A1.
XX
PD 22-JUN-2000.
XX
PF 15-DEC-1999; 99WO-US29826.
XX
PR 16-DEC-1998; 98US-0112558.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Falco SC, Kinney AJ;
XX
DR WPI: 2000-431595/37.
DR P-PSDB; AAY95314.
XX
PT Nucleic acids encoding plant CDP (cytosine diphosphate)-alcohol
PT phosphatidyltransferase polypeptide, useful for creating transgenic
PT plants in which the polypeptides are present at higher or lower levels
PT than normal -
XX
PS Claim 11; Page 46; 59pp; English.
XX
CC The present sequence is that of a contig of cDNA clones
CC p0005.cbmev33r and cs1.pk0038.g4 coding for corn
CC phosphatidylglycerophosphate synthase (see AAY95314), an enzyme
CC able to catalyze the formation of phosphatidylglycerophosphate,
CC the only phospholipid in chloroplast thylakoids. The cDNA clones
CC were isolated from corn immature ear and corn 5-week-old leaf
CC sheath cDNA libraries on the basis of the similarity of the
CC encoded enzyme with those from Arabidopsis thaliana and pseudomonas
CC fluorescens. The invention provides plant phospholipid biosynthetic
CC enzymes (see AAY95311-20) and the nucleic acids encoding them (see
CC AAA27928-37). The nucleic acids are useful for creating chimeric

CC genes encoding all or a portion of the phospholipid biosynthetic
CC chimeric, in sense or antisense orientation, where expression of the
CC chimeric gene results in production of altered levels of the enzyme
CC in a transformed host cell. Transgenic plants express the enzymes
CC at higher or lower levels than normal, or in cell types or
CC developmental stages in which they are not normally found.
CC Phospholipid biosynthetic enzymes expressed by host cells can be
CC used to identify inhibitors that may be useful as herbicides.
XX
SQ Sequence 1109 BP; 283 A; 247 C; 298 G; 280 T; 1 other;

Query Match 62.7%; Score 18.8; DB 21; Length 1109;
Best Local Similarity 76.7%; Pred. No. 2e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gccgtcttcacagtgctttatctgttcaact 30
||| ||||| || ||| ||||| |||
Db 530 GCTGCTTCCACATCCCTTAAATGTGTAAC 501

RESULT 15
AAH98348/C
ID AAH98348 standard; cDNA; 4965 BP.
XX
AC AAH98348;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 205.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM23689.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 319-321; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 4965 BP; 1683 A; 857 C; 971 G; 1454 T; 0 other;

Query Match 62.7%; Score 18.8; DB 22; Length 4965;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 tcttcacagttgctttatctgtt 26
||||| ||||| ||| |||
Db 2168 TCTTCCAGTTGCTTTTCTTTT 2147

Search completed: August 24, 2002, 22:11:11
Job time: 11921 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 18:16:45 ; Search time 204.64 Seconds
(without alignments)
36.010 Million cell updates/sec

Title: US-09-986-381-1
Perfect score: 30
Sequence: 1 gccgtcttcacgttgctttatctgttcaact 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28.4	94.7	600	1	US-08-472-239-2
2	18	60.0	1970	3	US-08-072-064-9
3	18	60.0	1970	5	PCT-US92-08558-3
4	18	60.0	2059	5	PCT-US92-08558-2
5	18	60.0	2066	3	US-08-072-064-2
6	18	60.0	2066	3	US-08-072-064-3
7	18	60.0	2066	3	US-08-072-064-5
8	18	60.0	2066	3	US-08-072-064-7
9	18	60.0	4906	2	US-09-047-026A-1
10	17.8	59.3	1170	4	US-09-419-459-3
11	17.8	59.3	1182	4	US-09-419-459-9
12	17.8	59.3	2110	4	US-09-419-459-1
13	17.6	58.7	309	2	US-08-633-879C-13
14	17.6	58.7	2104	1	US-08-592-126-96
15	17.6	58.7	2194	2	US-08-633-879C-3
16	17.4	58.0	1236	2	US-08-933-115-1
17	17.4	58.0	1236	4	US-09-205-008-1
18	17.2	57.3	1960	2	US-08-595-937A-1
19	17.2	57.3	1969	1	US-08-106-761-3
20	17.2	57.3	2457	1	US-08-486-013-68
21	17.2	57.3	2457	2	US-08-482-279-68
22	17.2	57.3	2457	2	US-08-342-268-68
23	17.2	57.3	2457	3	US-09-015-968-68
24	17.2	57.3	2457	4	US-09-397-386-68
25	17.2	57.3	2551	1	US-08-486-013-70
26	17.2	57.3	2551	2	US-08-482-279-70
27	17.2	57.3	2551	2	US-08-342-268-70

c 28	17.2	57.3	2551	3	US-09-015-968-70	Sequence 70, Appl
c 29	17.2	57.3	2551	4	US-09-397-386-70	Sequence 70, Appl
c 30	17	56.7	1133	4	US-08-630-915A-35	Sequence 35, Appl
c 31	17	56.7	1240	1	US-08-869-506-1	Sequence 1, Appli
c 32	17	56.7	1240	3	US-09-128-967-1	Sequence 1, Appli
c 33	17	56.7	2310	1	US-08-416-581B-2	Sequence 2, Appli
c 34	17	56.7	2787	1	US-08-416-581B-3	Sequence 3, Appli
c 35	17	56.7	2787	1	US-08-416-581B-4	Sequence 4, Appli
c 36	17	56.7	2787	3	US-09-288-461-1	Sequence 1, Appli
c 37	17	56.7	2787	4	US-09-087-465-5	Sequence 5, Appli
c 38	17	56.7	2800	2	US-08-874-138-1	Sequence 1, Appli
c 39	17	56.7	2800	2	US-08-874-138-5	Sequence 5, Appli
c 40	17	56.7	2800	4	US-08-879-941-1	Sequence 1, Appli
c 41	17	56.7	2800	4	US-08-879-941-3	Sequence 3, Appli
c 42	17	56.7	2800	4	US-09-747-116-1	Sequence 1, Appli
c 43	17	56.7	2800	4	US-09-747-116-3	Sequence 3, Appli
c 44	17	56.7	4519	4	US-08-976-259-6	Sequence 6, Appli
c 45	16.8	56.0	330	2	US-08-743-200-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-472-239-2
; Sequence 2, Application US/08472239
; Patent No. 5728526
; GENERAL INFORMATION:
; APPLICANT: GEORGE, JR., Albert L.
; APPLICANT: BHATNAGAR, Satish K.
; APPLICANT: NAZARENKO, Irena
; TITLE OF INVENTION: METHOD FOR ANALYZING A NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE AND KIT THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,239
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mukai, Robert G.
; REGISTRATION NUMBER: 28,531
; REFERENCE/DOCKET NUMBER: 020160-229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-472-239-2

Query Match 94.7%; Score 28.4; DB 1; Length 600;
Best Local Similarity 96.7%; Pred. No. 0.0022;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gccgtcttcacgttgctttatctgttcaact 30
||||| ||||||||| ||||||||| ||||||||| |||||||||

```

Db      33  CGCGTGTCCAGTTGCTTTATCTGTTCACT 62

RESULT      2
US-08-072-064-9/c
; Sequence 9, Application US/08072064
; Patent No. 6008046
; GENERAL INFORMATION:
; APPLICANT: FRENCH-CONSTANT, RICHARD H.
; APPLICANT: JACKSON, MEYER B.
; TITLE OF INVENTION: DRUG AND PESTICIDE SCREENING
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETER G. CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072.064
; FILING DATE: 19930602
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 770,881
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-00574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1970 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-072-064-9

Query Match      60.0%; Score 18; DB 3; Length 1970;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      2  ccgcttcaggctgtcttatctgttc 27
          ||||| | ||||| |||||
Db      1140  CGTCGAGCTGTGCTTTTCGTTC 1115

RESULT      3
PCT-US92-08558-3/c
; Sequence 3, Application PC/TUS9208558
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: MOLECULAR CLONING AND TRANSFORMATION OF CYCLODIENE RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US92-08558-2

Query Match 60.0%; Score 18; DB 5; Length 2059;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctccagttgctttatctgttc 27
||||| | ||||| |||||
Db 1229 CCGTCGAGCTGTGCTTTTCTGTC 1204

RESULT 5

US-08-072-064-2/c
; Sequence 2, Application US/08072064
; Patent No. 6008046

GENERAL INFORMATION:

APPLICANT: FRENCH-CONSTANT, RICHARD H.
APPLICANT: JACKSON, MEYER B.
TITLE OF INVENTION: DRUG AND PESTICIDE SCREENING
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETER G. CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,064
FILING DATE: 19930602
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 770,881
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPHD-00574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2066 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA

US-08-072-064-2

Query Match 60.0%; Score 18; DB 3; Length 2066;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctccagttgctttatctgttc 27
||||| | ||||| |||||
Db 1236 CCGTCGAGCTGTGCTTTTCTGTC 1211

RESULT 6

US-08-072-064-3/c
; Sequence 3, Application US/08072064
; Patent No. 6008046

GENERAL INFORMATION:
APPLICANT: FRENCH-CONSTANT, RICHARD H.
APPLICANT: JACKSON, MEYER B.
TITLE OF INVENTION: DRUG AND PESTICIDE SCREENING
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETER G. CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,064
FILING DATE: 19930602
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 770,881
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPHD-00574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 2066 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA

US-08-072-064-3

Query Match 60.0%; Score 18; DB 3; Length 2066;

Best Local Similarity 80.8%; Pred. No. 58;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctccagttgctttatctgttc 27
||||| | ||||| |||||
Db 1236 CCGTCGAGCTGTGCTTTTCTGTC 1211

RESULT 7

US-08-072-064-5/c
; Sequence 5, Application US/08072064
; Patent No. 6008046

GENERAL INFORMATION:

APPLICANT: FRENCH-CONSTANT, RICHARD H.
APPLICANT: JACKSON, MEYER B.
TITLE OF INVENTION: DRUG AND PESTICIDE SCREENING
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETER G. CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,064

```

; FILING DATE: 19930602
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 770,881
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPND-00574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2066 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-072-064-5.

Query Match 60.0%; Score 18; DB 3; Length 2066;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctccagtgctttatctgttc 27
    ||||| | ||||| |||||
DB 1236 CCGTCGAGCTGTTGCTTTTCTGTC 1211

RESULT 9
US-09-047-026A-1
; Sequence 1, Application US/09047026A
; Patent No. 5989897
; GENERAL INFORMATION:
; APPLICANT: Pillus, Lorraine
; APPLICANT: Clarke, Astrid
; APPLICANT: Lowell, Joanna
; APPLICANT: Jacobson, Sandra
; APPLICANT: Reifsnnyder, Cheryl
; TITLE OF INVENTION: Yeast Silencing Genes, Proteins and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/047.026A
; FILING DATE: 24-MAR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,375
; FILING DATE: 24-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 1-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4906 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1712..2728
; US-09-047-026A-1

Query Match 60.0%; Score 18; DB 2; Length 4906;
Best Local Similarity 80.8%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 tcttccagtgctttatctgttcact 30
    | ||| ||||| | |||||

```


Db 828 TTTTCTAGTGTCTTTTGTGTTTCTACT 853

RESULT 10

US-09-419-459-3
; Sequence 3, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1170)
US-09-419-459-3

Query Match 59.3%; Score 17.8; DB 4; Length 1170;
Best Local Similarity 75.9%; Pred. No. 65;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ccgtctccagtgctttatctgttccact 30
||||| ||||| | ||| ||||| |||||
Db 279 ccgtattccaactacttggtctggtcact 307

RESULT 11

US-09-419-459-9
; Sequence 9, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1182)
US-09-419-459-9

Query Match 59.3%; Score 17.8; DB 4; Length 1182;
Best Local Similarity 75.9%; Pred. No. 65;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ccgtctccagtgctttatctgttccact 30
||||| ||||| | ||| ||||| |||||
Db 291 ccgtattccaactacttggtctggtcact 319

RESULT 12

US-09-419-459-1

; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1

Query Match 59.3%; Score 17.8; DB 4; Length 2110;
Best Local Similarity 75.9%; Pred. No. 71;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ccgtctccagtgctttatctgttccact 30
||||| ||||| | ||| ||||| |||||
Db 608 ccgtattccaactacttggtctggtcact 636

RESULT 13

US-08-633-879C-13
; Sequence 13, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 3...239
OTHER INFORMATION:
US-08-633-879C-13

Query Match 58.7%; Score 17.6; DB 2; Length 309;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gccgtcttcacagttgctttatctg 24
|| ||||| ||||| ||||| |||||
Db 177 GCTGCTTCACAGTTGGTGATCTG 200

RESULT 14
US-08-592-126-96
; Sequence 96, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: G65.seq
US-08-592-126-96

Query Match 58.7%; Score 17.6; DB 1; Length 2104;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gccgtcttcacagttgctttatctg 24
|| ||||| ||||| ||||| |||||
Db 733 GCTGCTTCACAGTTGGTGATCTG 756

RESULT 15
US-08-633-879C-3
; Sequence 3, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2194 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 188...1792
; OTHER INFORMATION:
US-08-633-879C-3

Query Match 58.7%; Score 17.6; DB 2; Length 2194;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gccgtcttcacagttgctttatctg 24
|| ||||| ||||| ||||| |||||
Db 833 GCTGCTTCACAGTTGGTGATCTG 856

Search completed: August 24, 2002, 21:55:39
Job time: 13134 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 16:19:20 ; Search time 7654.23 Seconds
(without alignments)
52.900 Million cell updates/sec

Title: US-09-986-381-1

Perfect score: 30

Sequence: 1 gccgtcttcacagttgctttatctgttcaact 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21.6	72.0	543	10	BG467721
C 2	21.6	72.0	571	12	AQ461460 HS_5202_A
C 3	21.2	70.7	373	9	AA329627 EST33284
C 4	21.2	70.7	468	10	BE350494 ht14ei2.x
C 5	21.2	70.7	534	9	A1828798 tw70c02.x
C 6	21.2	70.7	534	10	BF805525 RC3-C1004
C 7	21.2	70.7	863	10	BF574189 602131348
C 8	21.2	70.7	2271	11	BC016881 Homo sapi
C 9	21	70.0	1083	12	CNS06RDW
C 10	20.8	69.3	672	10	BE470495 IPHDK0313
C 11	20.6	68.7	198	12	AQ348197 RPC111-12
C 12	20.6	68.7	348	10	BG003639 QV4-GN012
C 13	20.6	68.7	396	10	BG263068 WHE0946.G
C 14	20.6	68.7	447	10	BF929645
C 15	20.6	68.7	523	10	BF063348 7188e07.x
C 16	20.6	68.7	558	10	BM135631 WHE2622_B
C 17	20.6	68.7	602	12	B75397 RPC111-15N9

C 18	20.6	68.7	646	10	BI156388
C 19	20.6	68.7	670	10	BM171749
C 20	20.6	68.7	725	10	BG886654
C 21	20.6	68.7	734	10	BF529510
C 22	20.6	68.7	802	9	AU124625
C 23	20.6	68.7	816	9	AU117395
C 24	20.4	68.0	392	12	AQ241625
C 25	20.4	68.0	503	9	AV623600
C 26	20.4	68.0	596	9	A1734490
C 27	20.4	68.0	601	9	A1734538
C 28	20.4	68.0	794	12	BH140222
C 29	20	66.7	672	12	BH176734
C 30	20	66.7	672	12	CNS07JFV
C 31	20	66.7	926	10	BG256521
C 32	20	66.7	950	12	CNS041JP
C 33	19.8	66.0	421	9	AA484583
C 34	19.6	65.3	159	9	AA366528
C 35	19.6	65.3	296	9	BB268945
C 36	19.6	65.3	362	9	AA508802
C 37	19.6	65.3	363	9	AV737225
C 38	19.6	65.3	390	9	A1377706
C 39	19.6	65.3	391	9	A1289272
C 40	19.6	65.3	407	10	T935530
C 41	19.6	65.3	436	9	AA987277
C 42	19.6	65.3	470	9	AA972625
C 43	19.6	65.3	474	10	R92245
C 44	19.6	65.3	477	10	BF361438
C 45	19.6	65.3	499	10	N35561

ALIGNMENTS

RESULT 1

BG467721/c

LOCUS

DEFINITION

Na_L3_02C01_SAC Necator americanus (parasitic nematode) L3 Necator

americanus cDNA clone Na_L3_02C01 5' similar to O09015 MAX

INTERACTING PROTEIN 1 (MXIL PROTEIN) Rattus norvegicus, mRNA

sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG467721 543 bp mRNA linear EST 21-MAR-2001
Na_L3_02C01_SAC Necator americanus (parasitic nematode) L3 Necator
americanus cDNA clone Na_L3_02C01 5' similar to O09015 MAX
INTERACTING PROTEIN 1 (MXIL PROTEIN) Rattus norvegicus, mRNA
sequence.

BG467721 GI:13418325

EST.

Necator americanus.

Necator americanus

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;

Ancylostomatidae; Ancylostomatidae; Bunostominae; Necator.

1 (bases 1 to 543)

Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guillian,D., Hall

,N., Quayle,M. and Barrell,B.

Edinburgh University/Sanger Centre Nematode EST Project

Unpublished (2000)

Contact: Blaxter ML

Institute of Cell, Animal and Population Biology

University of Edinburgh

Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9

3JT, UK.

Tel: +44 131 650 6760

Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk

The library was prepared for Dr. David Pritchard University of

Nottingham Sequencing was performed by Claire Whitton ICAPB,

Edinburgh

PCR Primers

FORWARD: SAC

BACKWARD: 77PL

Plate: 02 row: C column: 01

Seq primer: SAC

High quality sequence stop: 486.

Location/Qualifiers

1..543

/organism="Necator americanus"

/db_xref="taxon:51031"

FEATURES

source

0; caps

SOURCE	ORGANISM	REFERENCE
human.	Homo sapiens	1 (bases 1 to 534)
	Eukaryota; Metazoa; Chordata	
	Mammalia; Eutheria; Primate	

TITLE	JOURNAL	MEDLINE	COMMENT
-------	---------	---------	---------

Simpson A.U.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

Email: hbe@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..198
 /organism="Homo sapiens"
 /db_xref="CDB:7548094"
 /db_xref="taxon:9606"
 /clone="RPCI-11-126D23"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"
 87 a 30 c 39 g 42 t

BASE COUNT
 ORIGIN

Query Match 68.7%; Score 20.6; DB 12; Length 198;
 Best Local Similarity 85.2%; Pred. No. 5.6e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 gtcttcagtgctttatctgttcttcaact 30
 ||||| ||||| ||||| ||||| ||
 Db 99 GTCTTCAGTGTCTTTTGTCTTCT 73

RESULT 12
 BG003639/C

LOCUS
 DEFINITION BG003639 348 bp mRNA linear EST 24-JAN-2001
 QV4-GN0122-191100-557-a10 GN0122 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG003639
 VERSION BG003639.1 GI:12444182
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 348)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zagoi,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV4st2-QV4-GN0122-191100-557-a10&t3=2000-11-19&t4=1>)
 Seg primer: puc.18 forward
 High quality sequence stop: 348.
 Location/Qualifiers
 1..348
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES
 source

ORIGIN

Query Match 68.7%; Score 20.6; DB 10; Length 396;
 Best Local Similarity 85.2%; Pred. No. 6.4e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcttcaggttgctttatctgttcac 28
 ||||| ||||| || ||||| ||||| ||
 Db 328 CCGTTTCCAGCTGATTATCTGTGGA 354

RESULT 14

BF929645/c 447 bp mRNA linear EST 19-JAN-2001
 LOCUS PM1-NT0235-081200-003-g12 NT0235 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF929645
 ACCESSION BF929645.1 GI:12327773
 VERSION
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 447)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

TITLE

Sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&st=PM1-NT0235-
 081200-003-g12&t3=2000-12-08&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 277.

FEATURES

source

1. .447
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0235"
 /dev_stage="Adult"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 135 a 112 c 114 g 86 t

ORIGIN

Query Match 68.7%; Score 20.6; DB 10; Length 447;
 Best Local Similarity 85.2%; Pred. No. 6.6e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtcttcaggttgctttatctgttcac 29
 ||||| ||||| ||||| ||||| ||
 Db 194 CGTCTCTCTGTGCTATCTGTTCCTCC 168

RESULT 15

LOCUS

BF063348

DEFINITION

7188e07.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:333076 3'-
 similar to SW:NCA2_HUMAN P13592 NEURAL CELL ADHESION MOLECULE,
 PHOSPHATIDYLINOSITOL-LINKED ISOFORM PRECURSOR ;, mRNA sequence.

ACCESSION

BF063348

VERSION

BF063348.1 GI:10822258

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 523)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Tumor Gene Index

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,

, Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL, send email to:

info@image.lnlni.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 465.

FEATURES

source

1. 523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:333076"

/clone_lib="NCI_CGAP_Col6"

/tissue_type="colon tumor, RER+"

/lab_host="DH10B"

/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Col6 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneIDs 1057416-1061255, and 1144584-1145351).

Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 91 a 138 c 136 g 158 t

ORIGIN

Query Match 68.7%; Score 20.6; DB 10; Length 523;

Best Local Similarity 85.2%; Pred. No. 6.8e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtcttcaggttgctttatctgttcac 29

||||| ||||| ||||| ||||| ||

Db 326 CGTCTCTCTGTGCTATCTGTTCCTCC 352

Search completed: August 24, 2002, 21:08:02

Job time: 17322 sec

• • •

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:52:24 ; Search time 2613.74 Seconds
(without alignments)
240.191 Million cell updates/sec

Title: US-09-986-381-2

Perfect score: 30

Sequence: 1 cctgatggcaaatgcccaattgcaggttaa 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pt.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vi.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

C 1	30	100.0	3407	9	HOMOTSP1	AF136270	Homo sapi
C 2	30	100.0	4233	9	HSM058JP1	AF135120	Homo sapi
C 3	30	100.0	4360	9	AB018045	AF018045	Homo sapi
C 4	30	100.0	20303	9	HSP53G	X54156	Human p53 (
C 5	30	100.0	20303	9	HSU94788	U94788	Human p53 (
C 6	30	100.0	159021	2	AC087388	AC087388	Homo sapi
C 7	30	100.0	160457	2	AC008049	AC008049	Homo sapi
C 8	21.6	72.0	189058	2	AC012101	AC012101	Homo sapi
C 9	20.6	68.7	72979	9	AL133259	AL133259	Human DNA
C 10	20.6	68.7	133997	9	AC015469	AC015469	Homo sapi
C 11	20.4	68.0	103054	2	AC014803	AC014803	Drosophil
C 12	20.4	68.0	181720	3	AC010114	AC010114	Drosophil
C 13	20.4	68.0	185435	2	AC067728	AC067728	Homo sapi
C 14	20.4	68.0	287018	3	AE003553	AE003553	Drosophil
C 15	20	66.7	1197	10	AF368860	AF368860	Rattus no
C 16	20	66.7	3279	3	DME291604	AJ291604	Drosophil
C 17	20	66.7	3642	3	AY060636	AY060636	Drosophil
C 18	20	66.7	55186	3	AC001655	AC001655	Drosophil
C 19	20	66.7	79420	2	AC017468	AC017468	Drosophil
C 20	20	66.7	110000	3	AE001572	AE001572	Drosophil
C 21	20	66.7	135214	9	AC000484	AC000484	Homo sapi
C 22	20	66.7	137143	9	AL451136	AL451136	Human DNA
C 23	20	66.7	149916	2	AL339540	AL339540	Homo sapi
C 24	20	66.7	152108	2	AC020282	AC020282	Drosophil
C 25	20	66.7	159513	3	AC010706	AC010706	Drosophil
C 26	20	66.7	165761	3	AC095016	AC095016	Drosophil
C 27	20	66.7	190642	3	AC095015	AC095015	Drosophil
C 28	20	66.7	301783	3	AE003499	AE003499	Drosophil
C 29	20	66.7	309357	3	AE003673	AE003673	Drosophil
C 30	19.8	66.0	29222	2	AC012917	AC012917	Drosophil
C 31	19.8	66.0	165098	3	AC018490	AC018490	Drosophil
C 32	19.8	66.0	169384	3	AC012164	AC012164	Drosophil
C 33	19.6	65.3	136624	2	AP004559	AP004559	Oryza sat
C 34	19.6	65.3	143409	9	AL162497	AL162497	human DNA
C 35	19.6	65.3	179132	10	AL358892	AL358892	Mouse DNA
C 36	19.6	65.3	180137	2	AC023081	AC023081	Homo sapi
C 37	19.6	65.3	184862	9	AC093905	AC093905	Homo sapi
C 38	19.6	65.3	190740	2	AC011864	AC011864	Homo sapi
C 39	19.6	65.3	191030	2	AL163545	AL163545	Homo sapi
C 40	19.6	65.3	205213	2	AC097537	AC097537	Homo sapi
C 41	19.6	65.3	215004	9	AC008507	AC008507	Homo sapi
C 42	19.4	64.7	1425	9	HS174D7F	HS174D7F	Chrysopa
C 43	19.4	64.7	1982	1	AF157621	AF157621	Chryseoba
C 44	19.4	64.7	14473	1	AE005377	AE005377	Escherich
C 45	19.4	64.7	80600	9	AL589941	AL589941	Human DNA

FEATURES		Location/Qualifiers	
source		1. .3407	
		/organism="Homo sapiens"	
		/db_xref="ATCC:CCL-185"	
		/db_xref="taxon:9606"	
		/cell_line="A549"	
		/note="A549 cells express wild-type p53 protein"	
exon		<216. .289	
		/gene="p53"	
exon		/number=2	
		407. .428	
		/gene="p53"	
exon		/number=3	
		522. .800	
		/gene="p53"	
exon		/number=4	
		1555. .1738	
		/gene="p53"	
exon		/number=5	
		1820. .1932	
		/gene="p53"	
exon		/number=6	
		2501. .2610	
		/gene="p53"	
exon		/number=7	
		2954. .3090	
		/gene="p53"	
exon		/number=8	
		3183. .3256	
		/gene="p53"	
		/number=9	
BASE COUNT	738 a	963 c	877 g 829 t
ORIGIN			
Query Match 100.0%; Score 30; DB 9; Length 3407;			
Best Local Similarity 100.0%; Pred. No. 0.0024;			
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	cctgatggcaaatgcccaattgcaggtaa 30	
Db	3364	CCTGATGGCAATGCCCAATTGCAGGTAA 3335	
RESULT 2			
LOCUS	HSM059JP1	3423 bp	DNA linear PRI 03-MAY-1999
DEFINITION	Homo sapiens tumor suppressor protein p53 (p53) gene, exons 2 through 9.		
ACCESSION	AF135120		
VERSION	AF135120.1	GI:4731629	
KEYWORDS	1		
SEGMENT	1 of 2		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Allalunis-Turner,M.J., Barton,G.M., Day,R.S. III, Dobler,K.D. and Mirzayans,R.		
TITLE	Isolation of two cell lines from a human malignant glioma specimen differing in sensitivity to radiation and chemotherapeutic drugs		
JOURNAL	Radiat. Res. 134 (3), 349-354 (1993)		
MEDLINE	93303270		
REFERENCE	2	(bases 1 to 3423)	
AUTHORS	Anderson,C.W., Kieleczawa,J. and Allalunis-Turner,M.J.		
TITLE	Human p53 from malignant glioma-derived M059J and M059K cells has a cancer-associated mutation in exon 8		
JOURNAL	Unpublished		
REFERENCE	3	(bases 1 to 3423)	
AUTHORS	Anderson,C.W., Kieleczawa,J. and Allalunis-Turner,M.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-MAR-1999) Biology, Brookhaven National Laboratory, 50		

FEATURES		Location/Qualifiers	
source		1. .3423	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/cell_type="M059J"	
		/note="M059J cells produce a mutant form of p53"	
exon		<216. .289	
		/gene="p53"	
exon		/number=2	
		407. .428	
		/gene="p53"	
exon		/number=3	
		538. .816	
		/gene="p53"	
exon		/number=4	
		1571. .1754	
		/gene="p53"	
exon		/number=5	
		1836. .1948	
		/gene="p53"	
exon		/number=6	
		2517. .2626	
		/gene="p53"	
exon		/number=7	
		2970. .3106	
		/gene="p53"	
variation		/number=8	
		2987	
		/gene="p53"	
		/replace="g"	
exon		3199. .3272	
		/gene="p53"	
		/number=9	
BASE COUNT	741 a	967 c	885 g 830 t
ORIGIN			
Query Match 100.0%; Score 30; DB 9; Length 3423;			
Best Local Similarity 100.0%; Pred. No. 0.0024;			
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	cctgatggcaaatgcccaattgcaggtaa 30	
Db	3380	CCTGATGGCAATGCCCAATTGCAGGTAA 3351	
RESULT 3			
LOCUS	AB018045/c		
DEFINITION	AB018045 Homo sapiens HSP70-1 gene for heat shock protein 72, spliced variant, partial cds.		
ACCESSION	AB018045		
VERSION	AB018045.1	GI:4691417	
KEYWORDS	HSP70-1; heat shock protein 72; HSP70-Hom; alternative splicing.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1	(sites)	
AUTHORS	Shimizu,S., Nomura,K., Ujihara,M. and Demura,H.		
TITLE	An additional exon of stress-inducible heat shock protein 70 gene (HSP70-1)		
JOURNAL	Biochem. Biophys. Res. Commun. 257 (1), 193-198 (1999)		
MEDLINE	99194576		
REFERENCE	2	(bases 1 to 4360)	
AUTHORS	Nomura,K. and Shimizu,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-SEP-1998) Kaoru Nomura, Tokyo Women's Medical University, Department of Medicine 2; 8-1 Kawadacho, Shinjyuku-ku, Tokyo 162-8666, Japan (E-mail:nomura7@parkcity.ne.jp, Tel:81-3-3353-8111(ex.39223), Fax:81-3-3357-6475) Sequence updated (26-Oct-1998).		
COMMENT			

```

FEATURES
  source      Location/Qualifiers
    1..4360
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="6"
      /map="6p21.3"
      /complement(1..196)
      /gene="HSP70-Hom"
      complement(1..196)
      /gene="HSP70-Hom"
      2323..2679
      /gene="HSP70-1"
      /note="alternative splicing"
      /number=1
      2323..4360
      /gene="HSP70-1"
      join(2632..2679,3955..>4360)
      /gene="HSP70-1"
      /note="spliced variant"
      /codon_start=1
      /product="heat shock protein 72"
      /protein_id="BAA77235.1"
      /db_xref="GI:4691418"
      /translation="MKHWPQVINDGDKPKVQVSYKGTAKFYPEIISWITKKKEI
      AEAYLGYPVNAVITYPAVFNDQSROATKDAGVIAGLVLRINEPTAAAIAYGLDRT
      GKGERNVLPFDLGGTDFDSVILTIIDGIFEVKATAGDTHLGGEDFDNRQ"
      3443..3954
      /gene="HSP70-1"
      /note="alternative splicing"
      /number=2
      3955..>4360
      /gene="HSP70-1"
      /note="alternative splicing"
      /number=3
      970 a 1242 c 1147 g 1001 t

BASE COUNT
ORIGIN

Query Match      100.0%; Score 30; DB 9; Length 4360;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cctgatgcaaatgcccaattgcaggtaa 30
|||||
Db 1845 CCTGATGCAAAATGCCCAATTGCAGGTAA 1816

/
RESULT 4
HSP53G/c
LOCUS
DEFINITION
  HSP53G 20303 bp DNA linear PRI 25-JUN-1997
  Human p53 gene for transformation related protein p53 (also called
  transformation-associated protein p53, cellular tumor antigen p53,
  and non-viral tumour antigen p53).
  X54156
  VERSION 1
  KEYWORDS anti-oncogene; cell cycle control; growth suppressor; heat shock
  protein 70; oncogene; p53 cellular tumour antigen; p53 gene;
  phosphoprotein; transforming capacity; tumor antigen.
  human.
SOURCE
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 20303)
  Chumakov,P.M.
  Direct Submission
  Submitted (02-AUG-1990) Chumakov P.M., Engelhardt Inst. of
  Molecular Biology, Academy of Science of the USSR, Vavilov St. 32,
  117984 Moscow, USSR
  2 (bases 1 to 20303)
  Chumakov,P.M., Almazov,V.P. and Jenkins,J.R.
  Unpublished
  3 (bases 1 to 20303)

AUTHORS Futreal,P.A., Barrett,J.C. and Wiseman,R.W.
TITLE An Alu polymorphism intragenic to the p53 gene
JOURNAL Nucleic Acids Res. 19 (24), 6977 (1991)
MEDLINE 92107726
COMMENT See also entries K03199, M14690, M14695, X01405, X02469, M22881-4,
M22887-8, M22894-8.
See also Mol. Cell. Biol. 6:1379-1385(1986);
and Mol. Cell. Biol. 7:961-963(1987).

FEATURES
  source      Location/Qualifiers
    1..20303
      /organism="Homo sapiens"
      /strain="caucasian"
      /db_xref="taxon:9606"
      /chromosome="17"
      /map="p13"
      join(843..949,11689..11790,11906..11927,12021..12299,
      13055..13238,13320..13432,14000..14109,14452..14588,
      14681..14754,17572..17678,18599..19876)
      /gene="p53"
      prim_transcript 843..19876
      exon 843..949
      /gene="p53"
      /number=1
      843..19876
      /gene="p53"
      950..11688
      /gene="p53"
      /number=1
      2581..2587
      /gene="p53"
      /note="5'-ALU flanking"
      2588..2877
      /gene="p53"
      /rpt_family="ALU"
      2890..2896
      /gene="p53"
      /note="3' ALU-flanking"
      3915..3929
      /gene="p53"
      /note="3' ALU-flanking"
      3950..4223
      /gene="p53"
      /rpt_family="ALU"
      4224..4238
      /gene="p53"
      /note="5'-ALU flanking"
      4319..4327
      /gene="p53"
      /note="5'-ALU flanking"
      4328..4603
      /gene="p53"
      /rpt_family="ALU"
      4631..4639
      /gene="p53"
      /note="3' ALU-flanking"
      4786..5574
      /gene="p53"
      /note="rearranged cluster"
      5802..5811
      /gene="p53"
      /note="5'-ALU flanking"
      5812..6100
      /gene="p53"
      /rpt_family="ALU"
      6127..6136
      /gene="p53"
      /note="3' ALU-flanking"
      6221..6236
      /gene="p53"
      /note="5'-ALU flanking"
      6237..6517

```

repeat_unit /gene="p53"
/rpt_family="ALU"
6531..6546
/gene="p53"
/note="3' ALU-flanking"
repeat_unit 6548..7812
/gene="p53"
/note="rearranged cluster"
repeat_unit 8703..8982
/rpt_family="ALU"
/gene="p53"
/rpt_family="ALU"
repeat_unit 9087..9098
/gene="p53"
/note="5'-ALU flanking"
repeat_unit 9099..9377
/gene="p53"
/rpt_family="ALU"
repeat_unit 9391..9402
/gene="p53"
/note="3' ALU-flanking"
repeat_unit 9513..10332
/gene="p53"
/note="rearranged cluster"
repeat_unit 11065..11069
/rpt_family="ALU"
/gene="p53"
/note="5'-ALU flanking"
repeat_unit 11070..11357
/gene="p53"
/rpt_family="ALU"
repeat_unit 11374..11378
/gene="p53"
/note="3' ALU-flanking"
exon 11689..11790
/gene="p53"
/number=2
CDS join(11717..11790,11906..11927,12021..12299,13055..13238,
13320..13432,14000..14109,14452..14588,14681..14754,
17572..17678,18599..18680)
/gene="p53"
/codon_start=1
/product="protein p53"
/protein_id="CRA38095.1"
/db_xref="GI:35214"
/translation="MEEQSDPSVEPPLSQETFSDLKLLPENNVLSPQLSQAMDLM
LSPDDEQFTDPGDEAPRPEAARVAPAPAPAPAPAPAPSPVPSQKT
YQSGYFLGFLHSGTAKSVCTYSPALNKMFCOLAKTCPVOLWVDSTPPCTRVRAM
AIVKQSHMTVVRCRPHHRCSDSDGLAPPOHLIRVSGNLEVEVLDNRNFRHSVVV
PYEPPEVGSDCCTIHYNYMCNSSCGMNRRLIILITLEDSSGNLLGRNFEVRVCA
CGRRRTTEENLRKKGPHHELPGSTKRALPNNTSSPPQKKKPLDGEYFTLIQIG
RERFERFRELNALELKDAQAGKEPGGSRANSHLSKKGOSTSRHKKLMFKTEGPD
S"
Intron 11791..11905
/gene="p53"
/number=2
exon 11906..11927
/gene="p53"
/number=3
Intron 11928..12020
/gene="p53"
/number=3
exon 12021..12299
/gene="p53"
/number=4
Intron 12300..13054
/gene="p53"
/number=4
repeat_unit 12588..12597
/gene="p53"
/note="5'-ALU flanking"
repeat_unit 12598..12882
/gene="p53"

repeat_unit /rpt_family="ALU"
12901..12910
/gene="p53"
/note="3' ALU-flanking"
exon 13055..13238
/gene="p53"
/number=5
Intron 13239..13319
/gene="p53"
/number=5
exon 13320..13432
/gene="p53"
/number=6
Intron 13433..13999
/gene="p53"
/number=6
repeat_unit 13617..13630
/gene="p53"
/note="5'-ALU flanking"
repeat_unit 13631..13913
/gene="p53"
/rpt_family="ALU"
repeat_unit 13930..13943
/gene="p53"
/note="3' ALU-flanking"
exon 14000..14109
/gene="p53"
/number=7
Intron 14110..14451
/gene="p53"
/number=7

Query Match 100.0%; Score 30; DB 9; Length 20303;
Best-Local Similarity 100.0%; Pred. No. 0.0026;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgcaggttaa 30
|||||
Db 14862 CCTGATGGCAATGCCCAATTGCAGGTAA 14833

FIGURE 3

RESULT 5
HSU94788/c
LOCUS HSU94788 20303 bp DNA linear PRI 17-JUL-2001
DEFINITION Human p53 (TP53) gene, complete cds.
ACCESSION U94788
VERSION U94788.1 GI:3041866
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 20303)
AUTHORS Rozenmuller,E.H., Kroyveld,A., Kreyveld,E., Leppers,F.G.J.,
Scheidel,K.C., Sliotweg,P.J. and Tilanus,M.G.J.
TITLE Sensitive detection of p53 mutation: analysis by direct sequencing
and multisequence analysis
JOURNAL Cancer Detect. Prev. 25 (2), 109-116 (2001)
MEDLINE 11341345
PUBMED 11341345
REFERENCE 2 (bases 1 to 20303)
AUTHORS Rozenmuller,E.H. and Tilanus,M.G.J.
TITLE P53 genomic sequence. Corrections and polymorphism
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 20303)
AUTHORS Rozenmuller,E.H.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-1997) Pathology, University Hospital Utrecht,
P.O.Box 85500, Utrecht 3508GA, The Netherlands
COMMENT This sequence describes corrections and polymorphisms in the
genomic p53 sequence and refers to the genomic TP53 sequence with
GenBank Accession Number X54156.
FEATURES Location/Qualifiers


```

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcaggtaa 30
|||||
Db 14862 CCTGATGGCAATGCCCAATTGCAGGTAA 14833

RESULT 6
AC087388/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-199F11 map 17, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
AC087388
AC087388.4 GI:16117576
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE
human.

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 159021)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Collangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galaean,J.,
Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Huime,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 13, 2001 this sequence version replaced gi:16041379.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11969
Center clone name: 199_F_11
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5822: contig of 5822 bp in length
* 5823 5922: gap of 100 bp
* 5923 43817: contig of 37895 bp in length
* 43818 43917: gap of 100 bp

```

```

* 43918 119706: contig of 75789 bp in length
* 119707 119806: gap of 100 bp
* 119807 153208: contig of 33402 bp in length
* 153209 153308: gap of 100 bp
* 153309 159021: contig of 5713 bp in length.
FEATURES
Location/Qualifiers
I..159021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-199F11"
/clone_lib="RP11-199F11 Human Male BAC"
BASE COUNT 40373 a 39760 c 38883 g 39366 t 639 others
ORIGIN

Query Match 100.0%; Score 30; DB 2; Length 159021;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcaggtaa 30
|||||
Db 80196 CCTGATGGCAATGCCCAATTGCAGGTAA 80167

RESULT 7
AC008049/c
LOCUS
DEFINITION
Homo sapiens clone RP11-199F11, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
AC008049
AC008049.37 GI:14787096
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE
human.

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 160457)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Bievat,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsif,P., Howard,S., Huber,J., Huylk,S., Hume,J.,
Joudah,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S.,
Oguy,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

```

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, F., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 160457)
Worley, K.C.
Submitted (16-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 17, 2001 this sequence version replaced gi:14328991.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMGV
Center clone name: RP11-199F11
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye: 66% of reads
Chemistry: Dye-terminator Big Dye: 66% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161924 bases at least Q40
Consensus quality: 169972 bases at least Q30
Consensus quality: 169776 bases at least Q20
Estimated insert size: 162538; sum-of-ctngs estimation
Quality coverage: 10.6x in Q20 bases; sum-of-ctngs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 55987: contig of 55987 bp in length
55988 56087: gap of unknown length
56088 96796: contig of 40709 bp in length
96797 96896: gap of unknown length
124682: contig of 27786 bp in length
124782: gap of unknown length
124783 136186: contig of 11404 bp in length
136187 136286: gap of unknown length
136287 141789: contig of 5503 bp in length
141790 141893: gap of unknown length
141890 149503: contig of 7614 bp in length
149504 149603: gap of unknown length
149604 154667: contig of 5064 bp in length
154668 154767: gap of unknown length
154768 156994: contig of 2227 bp in length
156995 157094: gap of unknown length
157095 160457: contig of 3363 bp in length.

FEATURES
source

1. 160457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-199F11"

BASE COUNT 39372 a 39284 c 40113 g 40841 t 847 others

ORIGIN

Query Match 100.0%; Score 30; DB 2; Length 160457;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgcaggttaa 30
|||||
Db 74525 CCTGATGGCAATGCCCCCAATTCAGGTAA 74496

RESULT 8

AC012101
LOCUS
DEFINITION
AC012101
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens chromosome 18 clone RP11-396D4 map 18, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
AC012101.6 GI:14971203
HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEPIN.
human.

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169058)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone RP11-396D4
Unpublished
2 (bases 1 to 169058)

REFERENCE

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boquslavsky, L., Bouqhalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 20, 2001 this sequence version replaced gi:14336464.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li014
Center clone name: 396_D_4

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 106897: contig of 106897 bp in length
106898 106997: gap of 100 bp
106998 169058: contig of 62061 bp in length.

FEATURES
source

1. 169058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-396D4"
/clone_lib="RPC1-11 Human Male BAC"

BASE COUNT ORIGIN	47668 a 34915 c 36862 g 49513 t 100 others	
Query Match	72.0%; Score 21.6; DB 2; Length 169058;	
Best Local Similarity	85.7%; Pred. No. 29;	
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		
QY 3 tgatggcaaatgcccaattgcaggttaa 30		
Db 40327 TGATTCAGAGGCCCATTCACAGGTAA 40354		
RESULT 9		
AL133259/C		
LOCUS	72979 bp DNA linear PRI 05-JUL-2000	
DEFINITION	Human DNA sequence from clone RP1-258E1 on chromosome 6 Contains part of the gene for CD83 antigen (activated B lymphocytes, immunoglobulin superfamily), ESTs, STSS and GSSs, complete sequence.	
ACCESSION	AL133259	
VERSION	AL133259.24 GI:8217427	
KEYWORDS	HTG; CD83.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 72979)	
TITLE	Tracey A.	
JOURNAL	Direct Submission	
COMMENT	Submitted (05-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequests@sanger.ac.uk On Jun 3, 2000 this sequence version replaced qi:7739056. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6 RP1-258E1 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTATOR: pCVPAC2 IMPORTANT: This sequence is not the entire insert of clone RP1-258E1 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP1-258E1 is at 72979 in this sequence. The true left end of clone RP11-359N11 is at 6773 in this sequence. The true right end of clone RP3-380E11 is at 100 in this sequence.	
FEATURES	Location/Qualifiers	
Source	1..72979	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="6"	
	/clone="RP1-258E1"	
	/clone_lib="RPCI-1"	
	810..981	
repeat_region		

repeat_region	/note="L2 repeat: matches 2178. .2349 of consensus"
	989. .1300
repeat_region	/note="Alusq repeat: matches 1. .311 of consensus"
	1654. .1685
repeat_region	/note="16 copies 2 mer aa 87% conserved"
	2690. .3137
repeat_region	/note="MLTIC repeat: matches 2. .464 of consensus"
	3274. .3333
repeat_region	/note="MER46A repeat: matches 178. .236 of consensus"
	3295. .3351
repeat_region	/note="MER46A repeat: matches 87. .146 of consensus"
	4274. .4452
repeat_region	/note="MIR repeat: matches 3. .184 of consensus"
	4887. .5191
misc_feature	/note="Alusx repeat: matches 1. .304 of consensus"
	complement(5954..6784)
misc_feature	/note="match: GSS: Em:AQ751861"
	6590. .7093
misc_feature	/note="match: GSS: Em:AQ732977"
	6782. .7372
repeat_region	/note="match: GSS: Em:AQ540540"
	7736. .7775
repeat_region	/note="20 copies 2 mer tg 90% conserved"
	8345. .8652
repeat_region	/note="AluJo repeat: matches 1. .300 of consensus"
	join(<9422..9650,11551..11657,13010..14820)
mRNA	/gene="CD83"
	/note="match: cDNAs: Em:AJ245551 Em:AF001036
	Em:AI815601 Em:AW403650 Em:N39825 Em:R19230
	Em:F07544 Em:R13302 Em:AW401678 Em:AA385462 Em:AA476799
	Em:AA391329 Em:AA490713 Em:AA083671 Em:AW404974
	Em:AF049583 Em:AI682302 Em:AW176128
	/product="dJ258E1.1 (CD83 antigen (activated B lymphocytes, immunoglobulin superfamily))"
	/evidence=not_experimental
gene	9422. .14820
	/gene="CD83"
CDS	join(<9422..9650,11551..11657,13010..13138)
	/gene="CD83"
	/note="match: proteins: Sw:Q01151 Tr:O88324"
	/codon_start=1
	/evidence=not_experimental
	/product="dJ258E1.1 (CD83 antigen (activated B lymphocytes, immunoglobulin superfamily))"
	/protein_id="CAB96723.1"
	/db_xref="GI:8953367"
	/translation="LLEGGERTPQEDHLRGOHVHKGQSGFDAPNPVSLKIR
	NTTSCSGYRCTLQDPDQGRNLGKVLRYTGCPAQKEETFKKYRAEIVLLALVI
	FYUULIFTKCFARLQSIIPDFSKAGMERAFLPVTSFNKHLGLVTPHKTELY"
	10192. .10237
repeat_region	/note="L2 repeat: matches 2701. .2744 of consensus"
	10357. .10585
repeat_region	/note="MIR repeat: matches 18. .254 of consensus"
	11267. .11307
repeat_region	/note="MIR repeat: matches 108. .148 of consensus"
	12557. .12649
repeat_region	/note="L2 repeat: matches 2592. .2685 of consensus"
	13139. .13538
misc_feature	/gene="CD83"
	/note="match: STS: Em:G11123"
	13758. .13763
polyA_signal	/gene="CD83"
	complement(14051..14576)
misc_feature	/note="match: STS: Em:D51194"
	complement(14297..14418)
misc_feature	/note="match: STS: Em:G43872"
	14559. .14564
polyA_signal	/gene="CD83"
	14622. .14759
repeat_region	/note="LIPAL3 repeat: matches 6018. .6156 of consensus"
	14759. .14764
polyA_signal	/gene="CD83"
	14776. .14781
polyA_signal	

```
/gene="CD83"
14820
/gene="CD83"
15172..15465
/note="AluY repeat: matches 1. .294 of consensus"
15584..15818
/note="MIR repeat: matches 9. .254 of consensus"
complement(15762..16559)
/note="match: GSS: Em:AQ745990"
16118..16410
/note="AluSg repeat: matches 1. .294 of consensus"
16411..16450
/note="20 copies 2 mer ag 80% conserved"
18127..18428
/note="AluY repeat: matches 3. .299 of consensus"
20334..20762
/note="MT1C repeat: matches 1. .464 of consensus"
21392..21477
/note="MIR repeat: matches 115. .207 of consensus"
22055..22372
/note="AluJb repeat: matches 1. .292 of consensus"
23809..23905
/note="MIR repeat: matches 49. .149 of consensus"
24897..24952
/note="28 copies 2 mer tg 80% conserved"
29460..29746
/note="AluSg repeat: matches 1. .291 of consensus"
29906..30007
/note="34 copies 3 mer tcc 76% conserved"
29940..30083
/note="3 copies 48 mer 75% conserved"
30029..30079
/note="17 copies 3 mer ctt 74% conserved"
31315..32143
/note="L2 repeat: matches 1571. .2383 of consensus"
32798..32871
/note="37 copies 2 mer tc 83% conserved"
33121..33213
/note="MT1G repeat: matches 8. .100 of consensus"
33368..33815
/note="L1ME repeat: matches 5307. .5762 of consensus"
34100..34357
/note="L1 repeat: matches 4738. .4997 of consensus"
34365..34415
/note="U7 repeat: matches 2. .53 of consensus"
34451..34771
/note="MER2 repeat: matches 1. .343 of consensus"
34933..35076
/note="3 copies 48 mer 85% conserved"
34945..35080
/note="68 copies 2 mer ct 60% conserved"
35083..35369
/note="AluJb repeat: matches 1. .297 of consensus"
35696..35877
/note="L2 repeat: matches 2578. .2750 of consensus"
36427..36707
/note="MSTA repeat: matches 1. .303 of consensus"
36712..37707
/note="MER1C repeat: matches 1. .1071 of consensus"
37715..37824
/note="MSTA repeat: matches 317. .426 of consensus"
complement(38622..39166)
/note="match: GSS: Em:AQ109375"
38685..38774
/note="MER20 repeat: matches 30. .119 of consensus"
39174..39539
/note="match: GSS: Em:B40528"
39346..39687
/note="AluSg repeat: matches 1. .310 of consensus"
```

Query Match 68.78; Score 20.6; DB 9; Length 72979;
Best Local Similarity 85.28; Pred.No. 83;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 3 tgatgcaaatgcccaatttcagta 29
|||||
Db 22888 TGTGTGAATYGCCCAATTCAGAA 22862
|||||

RESULT 10
AC015469 133997 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-IG10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC015469
ACCESSION AC015469
VERSION AC015469.3 GI:91119524
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 133997)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-IG10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 133997)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,B.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melárim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu.X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE Whitehead Institute/ MIT Center for Genome Research
JOURNAL Center code: WIBR
COMMENT Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2486
Center clone name: LG_10
-----
* NOTE: This record contains 146 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 852: contig of 852 bp in length
* 853 952: gap of 100 bp
* 953 1811: contig of 859 bp in length
* 1812 1911: gap of 100 bp
* 1912 2786: contig of 875 bp in length
* 2787 2886: gap of 100 bp
* 2887 3744: contig of 858 bp in length
* 3745 3844: gap of 100 bp
* 3845 4683: contig of 839 bp in length
* 4684 4783: gap of 100 bp
```

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2486
Center clone name: LG_10

* NOTE: This record contains 146 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 852: contig of 852 bp in length
* 853 952: gap of 100 bp
* 953 1811: contig of 859 bp in length
* 1812 1911: gap of 100 bp
* 1912 2786: contig of 875 bp in length
* 2787 2886: gap of 100 bp
* 2887 3744: contig of 858 bp in length
* 3745 3844: gap of 100 bp
* 3845 4683: contig of 839 bp in length
* 4684 4783: gap of 100 bp

```
* * 4784 5645: contig of 862 bp in length
* * 5646 5745: gap of 100 bp
* * 5746 6610: contig of 865 bp in length
* * 6611 6710: gap of 100 bp
* * 6711 7549: contig of 839 bp in length
* * 7550 7649: gap of 100 bp
* * 7650 8490: contig of 841 bp in length
* * 8491 8590: gap of 100 bp
* * 8591 9453: contig of 863 bp in length
* * 9454 9553: gap of 100 bp
* * 9554 10397: contig of 844 bp in length
* * 10398 10497: gap of 100 bp
* * 10498 11353: contig of 856 bp in length
* * 11354 11453: gap of 100 bp
* * 11454 12301: contig of 848 bp in length
* * 12302 12401: gap of 100 bp
* * 12402 13251: contig of 850 bp in length
* * 13252 13351: gap of 100 bp
* * 13352 14194: contig of 843 bp in length
* * 14195 14294: gap of 100 bp
* * 14295 15165: contig of 871 bp in length
* * 15166 15265: gap of 100 bp
* * 15266 16113: contig of 848 bp in length
* * 16114 16213: gap of 100 bp
* * 16214 17068: contig of 855 bp in length
* * 17069 17168: gap of 100 bp
* * 17169 18061: contig of 893 bp in length
* * 18062 18161: gap of 100 bp
* * 18162 19009: contig of 848 bp in length
* * 19010 19109: gap of 100 bp
* * 19110 19973: contig of 864 bp in length
* * 19974 20073: gap of 100 bp
* * 20074 20934: contig of 861 bp in length
* * 20935 21034: gap of 100 bp
* * 21035 21850: contig of 816 bp in length
* * 21851 21950: gap of 100 bp
* * 21951 22752: contig of 802 bp in length
* * 22753 22852: gap of 100 bp
* * 22853 23652: contig of 800 bp in length
* * 23653 23752: gap of 100 bp
* * 23753 24606: contig of 854 bp in length
* * 24607 24706: gap of 100 bp
* * 24707 25566: contig of 860 bp in length
* * 25567 25666: gap of 100 bp
* * 25667 26534: contig of 868 bp in length
* * 26535 26634: gap of 100 bp
* * 26635 27505: contig of 871 bp in length
* * 27506 27605: gap of 100 bp
* * 27606 28452: contig of 847 bp in length
* * 28453 28552: gap of 100 bp
* * 28553 29408: contig of 856 bp in length
* * 29409 29508: gap of 100 bp
* * 29509 30371: contig of 863 bp in length
* * 30372 30471: gap of 100 bp
* * 30472 31293: contig of 822 bp in length
* * 31294 31393: gap of 100 bp
* * 31394 32248: contig of 855 bp in length
* * 32249 32348: gap of 100 bp
* * 32349 33196: contig of 848 bp in length
* * 33197 33296: gap of 100 bp
* * 33297 34167: contig of 871 bp in length
* * 34168 34267: gap of 100 bp
* * 34268 35133: contig of 866 bp in length
* * 35134 35233: gap of 100 bp
* * 35234 36092: contig of 859 bp in length
* * 36093 36192: gap of 100 bp
* * 36193 37046: contig of 854 bp in length
* * 37047 37146: gap of 100 bp
* * 37147 38004: contig of 858 bp in length
* * 38005 38104: gap of 100 bp
* * 38105 38976: contig of 872 bp in length
* * 38977 39076: gap of 100 bp
* * 39077 39889: contig of 813 bp in length
```

```
* * 39890 39989: gap of 100 bp
* * 39990 40836: contig of 847 bp in length
* * 40837 40936: gap of 100 bp
* * 40937 41783: contig of 847 bp in length
* * 41784 41883: gap of 100 bp
* * 41884 42744: contig of 861 bp in length
* * 42745 42844: gap of 100 bp
* * 42845 43702: contig of 858 bp in length
* * 43703 43802: gap of 100 bp
* * 43803 44668: contig of 866 bp in length
* * 44669 44768: gap of 100 bp
* * 44769 45622: contig of 854 bp in length
* * 45623 45722: gap of 100 bp
* * 45723 46532: contig of 810 bp in length
* * 46533 46632: gap of 100 bp
* * 46633 47487: contig of 855 bp in length
* * 47488 47587: gap of 100 bp
* * 47588 48426: contig of 839 bp in length
* * 48427 48526: gap of 100 bp
* * 48527 49364: contig of 838 bp in length
* * 49365 49464: gap of 100 bp
* * 49465 50313: contig of 849 bp in length
* * 50314 50413: gap of 100 bp
* * 50414 51286: contig of 873 bp in length
* * 51287 51386: gap of 100 bp
* * 51387 52249: contig of 863 bp in length
* * 52250 52349: gap of 100 bp
* * 52350 53212: contig of 863 bp in length
* * 53213 53312: gap of 100 bp
* * 53313 54164: contig of 852 bp in length
* * 54165 54264: gap of 100 bp
* * 54265 55055: contig of 791 bp in length
* * 55056 55155: gap of 100 bp
* * 55156 55947: contig of 792 bp in length
* * 55948 56047: gap of 100 bp
* * 56048 56842: contig of 795 bp in length
* * 56843 56942: gap of 100 bp
* * 56943 57738: contig of 796 bp in length
* * 57739 57838: gap of 100 bp
* * 57839 58655: contig of 817 bp in length
* * 58656 58755: gap of 100 bp
* * 58756 59555: contig of 800 bp in length
* * 59556 59655: gap of 100 bp
* * 59656 60449: contig of 794 bp in length
* * 60450 60549: gap of 100 bp
* * 60550 61354: contig of 805 bp in length
* * 61355 61454: gap of 100 bp
* * 61455 62247: contig of 793 bp in length
* * 62248 62347: gap of 100 bp
* * 62348 63100: contig of 753 bp in length
* * 63101 63200: gap of 100 bp
* * 63201 63988: contig of 788 bp in length
* * 63989 64088: gap of 100 bp
* * 64089 64897: contig of 809 bp in length
* * 64898 64997: gap of 100 bp
* * 64998 65799: contig of 802 bp in length
* * 65800 65899: gap of 100 bp
* * 65900 66687: contig of 788 bp in length
* * 66688 66787: gap of 100 bp
* * 66788 67573: contig of 786 bp in length
* * 67574 67673: gap of 100 bp
* * 67674 68522: contig of 849 bp in length
```

Query Match 68.7%; Score 20.6; DB 2; Length 133997;
Best Local Similarity 85.2%; Pred. No. 85;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 taatgcaaatgcccaattgcagta 29
|||||
Db 96982 TGGTGAATAATGCCCAATTCAGGAA 97008

RESULT 11

JOURNAL

AC014803 103054 bp DNA linear HTG 16-NOV-1999
 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.

REFERENCE
AUTHORS

AC014803
 AC014803.1 GI:6436532
 HTG: HTGS_PHASE2.
 fruit fly.

ORGANISM

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 103054)

Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDM:10211121 by the submitter.
 For further information on this sequence e-mail to fly@celera.com.

* NOTE: this is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

Location/Qualifiers

1..103054

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

29118 a 22259 c 23349 g 29328 t

BASE COUNT

ORIGIN

Query Match 68.0% Score 20.4; DB 2; Length 103054;
 Best Local Similarity 80.0% Pred. No. 1 le+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgcagtaa 30

||| || ||||| ||||| ||||| |||

Db 84312 CCTTATAGCAAAATCCCAATGCAATTA 84341

RESULT 12

AC010114

LOCUS

Drosophila melanogaster 181720 bp DNA linear INV 03-FEB-2002
 Institute Drosophila BAC Library (Roswell Park Cancer

AC010114

AC010114.5 GI:15383757

HTG

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 181720)

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

Gocayne,J.D., Tabor,P., Williamson,A., Homsy,F.H.,

Dugan-Rocha,S.D., Sodergren,E.S., Hodson,A.H., Chen,R.C.,

Ayala,M.A., Scott,G.S., Worley,K.W., Anamatiades,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,

Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,

Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,

Draper,H., Emery-Cohen,A., Ferrier,S., Garg,N.D.S., Houck,J.,

Hostin,D., Howland,T.J., Hume,J., Ibegwan,C., Jalali,M., Kovar,C.,

Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,

Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Purl,V.,

Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,

Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 181720)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

Submitted (13-SEP-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
 Ogub,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshchar,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

REFERENCE
AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
 Ogub,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshchar,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

REFERENCE
AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
 Ogub,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshchar,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

REFERENCE
AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
 Ogub,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshchar,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE
JOURNAL

COMMENT
Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 31, 2001 this sequence version replaced gi:6996753.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

-STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES
source
1. .181720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3L"
/clone="RPC198-10P9"

BASE COUNT
ORIGIN
51148 a 39844 c 39470 g 51258 t

Query Match 68.0%; Score 20.4; DB 3; Length 181720;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccccaattgcaggtaa 30
||| || ||||| ||||| ||||| |||||

Db 106453 CCTTATAGCAAAATACCCCAATTGCAATTAA 106482

RESULT 13
AC067728

LOCUS
DEFINITION

AC067728 Homo sapiens chromosome 3 clone RP11-117F22, WORKING DRAFT SEQUENCE, 8 unordered pieces.

ACCESSION
VERSION

AC067728.15 GI:16303364

KEYWORDS
SOURCE

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

ORGANISM
human.

REFERENCE
AUTHORS

1 (bases 1 to 185435)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Benoit,J., Bimane,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,B., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Nwokenkwo,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE
JOURNAL

REFERENCE
AUTHORS

2 (bases 1 to 185435)
Worley,K.C.

TITLE
JOURNAL

Submitted (27-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 20, 2001 this sequence version replaced gi:16271879.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAXA
----- Summary Statistics


```
FEATURES             28049, Spain
Source              Location/Qualifiers
1..1197
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
10..96
/note="3' BP"
/codon_start=1
/product="3, non-translated beta-F1-ATPase mRNA-binding
protein"
/protein_id="AAK61874.1"
/db_xref="GI:14388973"
/translation="MKKHILLPLVLSLLMSSLQDSCGHEPS"

BASE COUNT          371 a 226 c 235 g 365 t
ORIGIN

Query Match          66.7%; Score 20; DB 10; Length 1197;
Best Local Similarity 82.1%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 ctgatggcaaatgcccaattgcaggta 29
|||||
Db 414 CTGATGCCAATGCAGAACTGAAGTA 441
```

Search completed: August 24, 2002, 21:53:09
Job time: 15474 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 22:11:11 ; Search time 905.65 seconds
(without alignments)
56.873 Million cell updates/sec

Title: US-09-986-381-2

Perfect score: 30

Sequence: 1 cctgattgcaaatgcccaattgcaggtaa 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21.6	72.0	160271	22 AAF85750	Bipolar affective
2	21.6	72.0	160271	22 AAF85756	Human chromosome 1
3	21.6	72.0	160271	22 AAS04858	Human chromosome 1
4	21.6	72.0	160271	22 AAS04864	Human chromosome 1
5	21.6	72.0	160271	22 AAS06667	Human chromosome 1
6	21.6	72.0	160271	22 AAH00997	160kb fragment of
7	21.6	72.0	160271	22 AAH23764	Human chromosome 1
8	21.6	72.0	160271	22 AAF85116	Nucleotide sequenc
9	20.4	68.0	2381	23 ABL22382	Drosophila melanog

C 10	20.4	68.0	4924	23 ABL09184	Drosophila melanog
C 11	20.4	68.0	4936	23 ABL20106	Drosophila melanog
C 12	20.4	66.7	2576	23 ABL10884	Drosophila melanog
C 13	20.4	66.7	3219	23 ABL14969	Drosophila melanog
C 14	20.4	66.7	5679	23 ABL14968	Drosophila melanog
C 15	19.0	63.3	2375	18 AAV74458	Staphylococcus aur
C 16	19.0	63.3	7709	22 AAL05397	Human reproductive
C 17	18.8	62.7	891	22 AAH1386	Escherichia coli p
C 18	18.8	62.7	1101	23 AAS81969	DNA encoding novel
C 19	18.8	62.7	1101	23 AAS93718	DNA encoding novel
C 20	18.8	62.7	1534	23 AAS89112	DNA encoding novel
C 21	18.8	62.7	1583	23 AAS69369	DNA encoding novel
C 22	18.8	62.7	7816	22 AAK52934	Human polynucleoti
C 23	18.8	62.7	7818	22 AAK51950	Human polynucleoti
C 24	18.8	62.7	14533	23 AAS65883	DNA encoding novel
C 25	18.6	62.0	1037	22 AAH48630	Human MLP promoter
C 26	18.6	62.0	3950	20 AAX25521	DNA coding for mod
C 27	18.4	61.3	369	22 AAH81731	Human differential
C 28	18.4	61.3	1968	19 AAV52252	Streptococcus pneu
C 29	18.4	61.3	6710	23 ABL02156	Drosophila melanog
C 30	18.2	60.7	355	21 AAC28593	Human secreted pro
C 31	18.0	60.0	942	22 AAH32367	Human olfactory re
C 32	18.0	60.0	945	22 AAS42397	Human cDNA encodin
C 33	18.0	60.0	1118	21 AAC49190	Arabidopsis thalia
C 34	18.0	60.0	1120	21 AAC35386	Arabidopsis thalia
C 35	18.0	60.0	1416	23 AAS56293	Salmonella typhi D
C 36	18.0	60.0	1824	23 AAS71981	DNA encoding novel
C 37	18.0	60.0	2265	23 AAS88612	DNA encoding novel
C 38	18.0	60.0	23802	22 AAK75632	Human immune/haema
C 39	17.8	59.3	300	21 AAA01318	Human colon cancer
C 40	17.8	59.3	542	22 AAH11419	Human cDNA clone (
C 41	17.8	59.3	974	20 AAZ25530	pGR2PL6 cDNA seque
C 42	17.8	59.3	975	19 AAV29855	Streptococcus pneu
C 43	17.8	59.3	975	22 AAD06210	Streptococcus pneu
C 44	17.8	59.3	975	22 AAC82127	S. pneumoniae FabH
C 45	17.8	59.3	1007	20 AAZ25531	pGR2PL3 cDNA seque

ALIGNMENTS

RESULT 1

AAF85750

ID AAF85750 standard; DNA; 160271 BP.

XX

AC AAF85750;

XX

DT 10-DEC-2001 (first entry)

XX

DE Bipolar affective disorder associated gene fsh24 genomic sequence.

XX

XX Human; bipolar affective disorder; BAD; fsh24; neuropsychiatric disorder;

KW manic depression; schizophrenia; attention deficit disorder;

KW schizoaffective disorder; neurodegenerative disorder; gene therapy;

KW chromosome 18q; ds.

XX Homo sapiens.

OS

PN WO200134173-A1.

XX

PD 17-MAY-2001.

XX

PF 07-NOV-2000; 2000WO-US30611.

XX

PR 08-NOV-1999; 99US-0164041.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX (REGC) UNIV CALIFORNIA.

PI

XX Chen H, Freimer NB;

DR WPI; 2001-316487/33.

XX

PT Nucleic acids encoding fsh24 polypeptides, useful for preventing,
PT diagnosing and treating neuropsychiatric disorders, e.g. schizophrenia,
PT attention deficit disorder, bipolar affective disorders and unipolar
PT disorders -
XX
XX
PS Claim 1; Fig 1B; 155pp; English.
XX
XX The present invention provides the gene and cDNA sequences of the human
CC fsh24 gene. This is found on human chromosome 18q and is associated with
CC neuropsychiatric disorders such as schizophrenia, attention deficit
CC disorder, schizoaffective disorders, bipolar and unipolar affective
CC disorders, dysthymic disorder, major depressive disorder, mania,
CC obsessive-compulsive disorder, psychoactive substance use disorders,
CC anxiety and panic disorder, and neurodegenerative disorders such as
CC Parkinson's, Alzheimer's and Huntington's diseases, senile dementia,
CC amyotrophic lateral sclerosis and Gilles de la Tourette's Syndrome,
CC hypertension and sleep disorders. The sequences can be used in the
CC diagnosis and treatment of these conditions. The present sequence is the
CC genomic sequence of the invention.
XX
SQ Sequence 160271 BP; 45619 A; 32964 C; 34928 G; 46702 T; 58 other;

Query Match 72.0%; Score 21.6; DB 22; Length 160271;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 3 tgatggcaaatgcccaattgcaggtaa 30
||||| ||||| ||||| || |||||
Db 12174 tgattgcaagggcccaattccaggtaa 12201

RESULT 2
AAF85756
ID AAF85756 standard; DNA; 160271 BP.
XX
AC AAF85756;
XX
XX 10-DEC-2001 (first entry)
XX
XX Human chromosome 18q bipolar affective disorder related fragment.
DE
DE
XX
XX Human; bipolar affective disorder; BAD; neuropsychiatric disorder;
KW manic depression; schizophrenia; attention deficit disorder;
KW schizoaffective disorder; neurodegenerative disorder; gene therapy;
KW chromosome 18q; ds.
XX
XX Homo sapiens.
OS
XX
PN WO200134771-A2.
XX
PD 17-MAY-2001.
XX
XX 07-NOV-2000; 2000WO-US30637.
PF
XX
XX 08-NOV-1999; 99US-01639973.
PR
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA (REGC) UNIV CALIFORNIA.
PA
XX
XX Chen H, Freimer NB;
PI
XX
XX WPI; 2001-316493/33.
DR
XX
XX Polynucleotide sequences from the long arm of human chromosome 18 are
PT used in diagnosis of neuropsychiatric disorders such as schizophrenia
PT and bipolar affective disorders and for detecting compounds suitable
PT for treating such disorders -
XX
XX Claim 1; Fig 1B; 175pp; English.
PS
XX
XX The present invention provides a nucleotide sequence found on human
CC chromosome 18q between the bipolar affective disorder (BAD) related

CC markers BAD18ct22 and BAD18cag1. The region is also associated with
CC neuropsychiatric disorders such as schizophrenia, attention deficit
CC disorder, schizoaffective disorders, bipolar and unipolar affective
CC disorders, dysthymic disorder, major depressive disorder, mania,
CC obsessive-compulsive disorder, psychoactive substance use disorders,
CC anxiety and panic disorder, and neurodegenerative disorders such as
CC Parkinson's, Alzheimer's and Huntington's diseases, senile dementia,
CC amyotrophic lateral sclerosis and Gilles de la Tourette's Syndrome,
CC hypertension and sleep disorders. The sequence can be used in the
CC diagnosis and treatment of these conditions. The present sequence the
CC nucleotide sequence of the invention.
XX
SQ Sequence 160271 BP; 45619 A; 32964 C; 34928 G; 46702 T; 58 other;

Query Match 72.0%; Score 21.6; DB 22; Length 160271;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 3 tgatggcaaatgcccaattgcaggtaa 30
||||| ||||| ||||| || |||||
Db 12174 tgattgcaagggcccaattccaggtaa 12201

RESULT 3
AAS04858
ID AAS04858 standard; DNA; 160271 BP.
XX
AC AAS04858;
XX
XX 07-SEP-2001 (first entry)
XX
XX Human chromosome 18, BAD18ct22-BAD18cag1.
DE
DE
XX
XX Human; chr18q: fsh30; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; antibody; schizophrenia; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW ds.
OS
XX Homo sapiens.
XX
PH Key Location/Qualifiers
FT misc_feature 28441..144419
FT /tag= a
FT /note= "Interval associated with neuropsychiatric
FT disorders"
FT 28441..29265
FT /tag= b
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT primer_bind 28384..28405
FT /tag= c
FT /label= BADct22_forward
FT 28547..28572
FT /tag= d
FT /rpt_type= TANDEM
FT /note= "BADct22 marker"
FT primer_bind complement (28547..28572)
FT /tag= e
FT /label= BADct2_reverse
FT 29683..39587
FT /tag= f
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 40284..43253
FT /tag= g
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 43518..46075
FT /tag= h
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 43518..46075
FT misc_feature

FT FT /tag= i
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 47264..52284
FT FT /tag= j
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 52672..56935
FT FT /tag= k
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 57032..57726
FT FT /tag= l
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 58065..59057
FT FT /tag= m
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 59815..60471
FT FT /tag= n
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 60870..62451
FT FT /tag= o
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 62543..63268
FT FT /tag= p
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 63494..66959
FT FT /tag= q
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 67964..69670
FT FT /tag= r
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 70643..70749
FT FT /tag= s
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 71051..72295
FT FT /tag= t
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 72858..76408
FT FT /tag= u
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 76797..77123
FT FT /tag= v
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 77663..78170
FT FT /tag= w
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 78463..80173
FT FT /tag= x
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 80466..81519
FT FT /tag= y
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 81888..85946
FT FT /tag= z
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 86346..87569
FT FT /tag= aa

FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 87461..87958
FT FT /tag= ab
FT FT /product= fsh30_gene_product
FT FT /note= "This sequence is specifically claimed"
FT FT 88674..89188
FT FT /tag= ac
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 89459..89745
FT FT /tag= ad
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 90436..92299
FT FT /tag= ae
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 92406..94789
FT FT /tag= af
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 93556..100121
FT FT /tag= ag
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 100530..101382
FT FT /tag= ah
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 101798..103865
FT FT /tag= ai
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 101798..103865
FT FT /tag= aj
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 104486..109841
FT FT /tag= ak
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 109953..110561
FT FT /tag= al
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 111000..113482
FT FT /tag= am
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 113774..116253
FT FT /tag= an
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 116846..117907
FT FT /tag= ao
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 117999..118623
FT FT /tag= ap
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 118865..122881
FT FT /tag= aq
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed"
FT FT 122978..186088
FT FT /tag= ar
FT FT /label= EST_matching_region
FT FT /note= "This sequence is specifically claimed. The
FT FT range given for this feature is as stated in the
FT FT specification but is clearly wrong"
FT FT 129508..130413
FT FT misc_feature

```

FT      /*tag= as
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 131138..134228
FT      /*tag= at
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"

Query Match      72.0%; Score 21.6; DB 22; Length 160271;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      3 tgatggcaaatgcccgaattgcaggtaa 30
      |||| |||| |||| |||| |||| ||||
Db 12174 tgattgcaagggcccaatttccaggtaa 12201

RESULT 4
AAS04864
ID AAS04864 standard; DNA; 160271 BP.
XX AC
AAS04864;
XX AC
XX 07-SEP-2001 (first entry)
DT
XX
XX Human chromosome 18, BAD18ct22-BAD18cag1.
XX
XX Human; chr18q; fsh27; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; antibody; schizophrenia; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH misc_feature 28441..144419
FT      /*tag= a
FT      /note= "Interval associated with neuropsychiatric
FT      disorders"
FT      misc_feature 28441..29265
FT      /*tag= b
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      primer_bind 28384..28405
FT      /*tag= c
FT      /label= BADct22_forward
FT      28547..28572
FT      /*tag= d
FT      /rpt_type= TANDEM
FT      /note= "BADct22 marker"
FT      primer_bind complement (28547..28572)
FT      /*tag= e
FT      /label= BADct2_reverse
FT      29683..39587
FT      /*tag= f
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 40284..43253
FT      /*tag= g
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 43518..46075
FT      /*tag= h
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 43518..46075
FT      /*tag= i
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 47264..52284
FT      /*tag= j
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"

```

```

FT      misc_feature 52672..56935
FT      /*tag= k
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 57032..57726
FT      /*tag= l
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 58065..59057
FT      /*tag= m
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 59815..60471
FT      /*tag= n
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 60870..62451
FT      /*tag= o
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 62543..63268
FT      /*tag= p
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 63494..66959
FT      /*tag= q
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 67964..69670
FT      /*tag= r
FT      /label= EST_matching_area
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 70643..70749
FT      /*tag= s
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 71051..72295
FT      /*tag= t
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 72858..76408
FT      /*tag= u
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 76797..77123
FT      /*tag= v
FT      /label= EST_matching_area
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 77663..78170
FT      /*tag= w
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 78463..80173
FT      /*tag= x
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 80466..81519
FT      /*tag= y
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 81888..85946
FT      /*tag= z
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 86346..87569
FT      /*tag= aa
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 88674..89188
FT      /*tag= ab
FT      /label= EST_matching_region
FT      /note= "This sequence is specifically claimed"
FT      misc_feature 89459..89745

```



```
FT /*tag= ac
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 90436..92299
FT /*tag= ad
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 92406..94789
FT /*tag= ae
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 95536..100121
FT /*tag= af
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 100530..101382
FT /*tag= ag
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 101798..103865
FT /*tag= ah
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 101798..103865
FT /*tag= ai
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 104486..109841
FT /*tag= aj
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 109953..110561
FT /*tag= ak
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 111000..113482
FT /*tag= al
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 113774..116253
FT /*tag= am
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 115814..119863
FT /*tag= an
FT /product= fsh27_gene.product
FT /note= "This sequence is specifically claimed"
FT 116846..117907
FT /*tag= ao
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 117999..118623
FT /*tag= ap
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 118865..122881
FT /*tag= aq
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 122978..186088
FT /*tag= ar
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed. The
FT range given for this feature is as stated in the
FT specification but is clearly wrong"
FT 129508..130413
FT /*tag= as
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
FT 131138..134228
FT /*tag= at
FT /label= EST_matching_region
FT /note= "This sequence is specifically claimed"
```

```
Query Match 72.0%; Score 21.6; DB 22; Length 160271;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 tgatggcaaatgcccaattgcaggtaa 30
   ||| |||| |||| || |||||
Db 12174 tgattgcaaaagcccccatttccaggtaa 12201

RESULT 5
AAS06667
ID AAS06667 standard; DNA; 160271 BP.
XX AC AAS06667;
XX 12-SEP-2001 (first entry)
XX DE Human chromosome 18q, 160kb sequence.
XX KW Human; 18q; fsh23; neuropsychiatric disorder; schizophrenia;
XX KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
XX KW brain tumour; diabetes; angina pectoris; ds.
XX OS Homo sapiens.
XX FH Location/Qualifiers
FT primer_bind
FT 28384..28405
FT /*tag= a
FT /note= "PCR primer BAD18ct22 forward"
FT 28441..29265
FT /*tag= b
FT /note= "This region is specifically claimed"
FT 29683..39587
FT /*tag= c
FT /note= "This region is specifically claimed"
FT 28441..144419
FT /*tag= d
FT /note= "Region associated with neuropsychiatric
FT disorders"
FT primer_bind
FT complement (28547..28572)
FT /*tag= e
FT /note= "PCR primer BAD18ct22 reverse"
FT 40284..43253
FT /*tag= f
FT /note= "This region is specifically claimed"
FT 43518..46075
FT /*tag= g
FT /note= "This region is specifically claimed"
FT 47264..52284
FT /*tag= h
FT /note= "This region is specifically claimed"
FT 51458..63491
FT /*tag= i
FT /product= "Fsh23"
FT /note= "Fsh23 gene region"
FT 52672..56935
FT /*tag= j
FT /note= "This region is specifically claimed"
FT 57032..57726
FT /*tag= k
FT /note= "This region is specifically claimed"
FT 58065..59057
FT /*tag= l
FT /note= "This region is specifically claimed"
FT 59815..60471
FT /*tag= m
FT /note= "This region is specifically claimed"
FT 60870..62451
FT /*tag= n
FT /note= "This region is specifically claimed"
FT 62543..63268
FT /*tag= o
```

FT misc_feature /note= "This region is specifically claimed"
63494..66969
/*tag= p
FT misc_feature /note= "This region is specifically claimed"
67964..69670
/*tag= q
FT misc_feature /note= "This region is specifically claimed"
70643..70749
/*tag= r
FT misc_feature /note= "This region is specifically claimed"
71051..72295
/*tag= s
FT misc_feature /note= "This region is specifically claimed"
72858..76408
/*tag= t
FT misc_feature /note= "This region is specifically claimed"
76797..77123
/*tag= u
FT misc_feature /note= "This region is specifically claimed"
77663..78170
/*tag= v
FT misc_feature /note= "This region is specifically claimed"
78463..80173
/*tag= w
FT misc_feature /note= "This region is specifically claimed"
80466..81519
/*tag= x
FT misc_feature /note= "This region is specifically claimed"
81888..85946
/*tag= y
FT misc_feature /note= "This region is specifically claimed"
86346..87569
/*tag= z
FT misc_feature /note= "This region is specifically claimed"
88674..89188
/*tag= aa
FT misc_feature /note= "This region is specifically claimed"
89459..89745
/*tag= ab
FT misc_feature /note= "This region is specifically claimed"
90436..92299
/*tag= ac
FT misc_feature /note= "This region is specifically claimed"
92406..94789
/*tag= ad
FT misc_feature /note= "This region is specifically claimed"
95556..100121
/*tag= ae
FT misc_feature /note= "This region is specifically claimed"
100530..101382
/*tag= af
FT misc_feature /note= "This region is specifically claimed"
101798..103865
/*tag= ag
FT misc_feature /note= "This region is specifically claimed"
104480..109841
/*tag= ah
FT misc_feature /note= "This region is specifically claimed"
109953..110561
/*tag= ai
FT misc_feature /note= "This region is specifically claimed"
11000..113482
/*tag= aj
FT misc_feature /note= "This region is specifically claimed"
113774..116253
/*tag= ak
FT misc_feature /note= "This region is specifically claimed"
116846..117907
/*tag= al
FT misc_feature /note= "This region is specifically claimed"
117999..118623
/*tag= am
FT misc_feature /note= "This region is specifically claimed"

FT misc_feature 118865..122881
/*tag= an
FT misc_feature /note= "This region is specifically claimed"
122978..186088
/*tag= ao
FT misc_feature /note= "This region is specifically claimed.
This is an erroneous range given that the
sequence is only 160kb in size"
129508..130413
/*tag= ap
FT misc_feature /note= "This region is specifically claimed"
131138..134228
/*tag= aq
FT misc_feature /note= "This region is specifically claimed"
134517..135473
/*tag= ar
FT misc_feature /note= "This region is specifically claimed"
135815..139983
/*tag= as
FT misc_feature /note= "This region is specifically claimed"
140683..144419
/*tag= at
FT primer_bind /note= "This region is specifically claimed"
144302..144325
/*tag= au
FT primer_bind /note= "PCR primer BAD18cag1 forward"
complement (144477..144501)
/*tag= av
FT primer_bind /note= "PCR primer BAD18cag1 reverse"
144302..144325
/*tag= au
FT primer_bind /note= "PCR primer BAD18cag1 forward"
complement (144477..144501)
/*tag= av
FT primer_bind /note= "PCR primer BAD18cag1 reverse"
144302..144325
/*tag= au
XX
PN WO200134772-A2.
XX 17-MAY-2001.
XX 08-NOV-2000; 2000WO-US30819.
XX 08-NOV-1999; 99US-0164042.
XX (MILL-) MILLENNIUM PHARM INC.
(REGC) UNIV CALIFORNIA.
XX Chen H, Freimer NB;
XX WPI; 2001-343601/36.
DR Novel mammalian fsh23 polynucleotide for diagnostic evaluation, genetic
XX testing and prognosis of fsh23-related disorders such as
PT neuropsychiatric disorders including schizophrenia, bipolar affective
PT disorder -
XX
PS Claim 1; Fig 1B; 198pp; English.
XX The sequence represents 160kb of human chromosome 18q containing
CC fsh23 gene, located in a region associated with neuropsychiatric
CC disorders. The fsh23 gene and polypeptide, its fragment, analog or
CC mimetic is useful for treating a fsh23-related disorder or fsh23-mediated
CC process such as neuropsychiatric disorders e.g., schizophrenia,
CC attention deficit disorder, schizoaffective disorder, bipolar affective
CC disorder and unipolar disorder. A cell harbouring the gene is engineered
CC ex vivo to express an unimpaired fsh23 protein. The gene and protein are
CC also useful for treating neurodegenerative disorders such as Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's disease, and Gilles de la Tourette's syndrome, and
CC hypertension and sleep disorders, mania, obsessive-compulsive disorder,
CC and anxiety. The gene and protein are also useful for the diagnostic
CC evaluation, genetic testing and/or prognosis of a fsh23-related disorder,
CC
Query Match 72.0%; Score 21.6; DB 22; Length 160271;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 3 tgatggcaaatgcccaattgcaggttaa 30
||||| ||||| ||||| || |||||

Db 12174 tgattgcaaggcccccatttcaggtaa 12201

RESULT 6

AAH40997

ID AAH40997 standard; DNA; 160271 BP.

XX AC AAH40997;

XX 21-AUG-2001 (first entry)

XX DT

XX DE 160kb fragment of the long arm of human chromosome 18q.

XX KW Chromosome 18q; fsh25; human; neuropsychiatric disorder; schizophrenia; attention deficit disorder; schizoaffective disorder; BAD; mania; bipolar affective disorder; unipolar disorder; depression; ds; BP; bipolar mood disorder.

XX OS Homo sapiens.

XX FH Key

FT primer_bind

FT complement (28384..28405)

FT /*tag= a

FT /*note= "BADct22 specific primer"

FT 28440..28483

FT /*tag= b

FT /*note= "BADct22 marker sequence"

FT 28547..28572

FT /*tag= c

FT /*note= "BADct22 specific primer"

FT complement (116254..116523)

FT /*tag= d

FT /*product= "fsh25 protein"

FT complement (144302..144325)

FT /*tag= e

FT /*note= "BAD18cag1 specific primer"

FT 144388..144420

FT /*tag= f

FT /*note= "BAD18cag1 marker sequence"

FT 144477..144501

FT /*tag= g

FT /*note= "BAD18cag1 specific primer"

XX WO200133965-A1.

XX 17-MAY-2001.

XX 07-NOV-2000; 2000WO-US30636.

XX 08-NOV-1999; 990S-0164038.

XX (MILL-) MILLENNIUM PHARM INC.

XX (REGC) UNIV CALIFORNIA.

XX Chen H, Freimer NB;

XX WPI; 2001-308772/32.

XX Nucleic acids encoding mammalian fsh25 polypeptides associated with neuropsychiatric disorders, useful for treating e.g. bipolar affective disorders -

XX Claim 1; Fig 1B; 153pp; English.

XX This invention relates to the present sequence which represents a 160kb fragment of the long arm of human chromosome 18q. The sequence includes the 18q interval associated with neuropsychiatric disorders (nucleotides 28441-144419). Included in this region is the fsh25 gene, which is involved in neuropsychiatric disorders. Included in the invention are fragments of the DNA sequence, and antibodies which bind to its protein products. Fsh25 DNA and protein sequences, a vector containing the DNA sequence, and an antibody directed against the protein product may be used in the prevention, diagnosis and treatment of diseases associated

CC with inappropriate fsh25 expression. Disorders that may be prevented, diagnosed and/or treated using the DNA, protein, vector and antibody include neuropsychiatric disorders, such as schizophrenia, attention deficit disorder, a schizoaffective disorder, a bipolar affective disorder or a unipolar disorder.

CC SQ Sequence 160271 BP; 45619 A; 32965 C; 34927 G; 46702 T; 58 other;

Query Match 72.0%; Score 21.6; DB 22; Length 160271;

Best Local Similarity 85.7%; Pred. No. 15;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 tgatggcaatgcccatttcaggtaa 30

Db 12174 tgattgcaaggcccccatttcaggtaa 12201

RESULT 7

AAH23764

ID AAH23764 standard; DNA; 160271 BP.

XX AC AAH23764;

XX DT 13-AUG-2001 (first entry)

XX DE Human chromosome 18q interval containing the fsh26 gene.

XX KW Human; chromosome 18q; fsh26; nootropic; neuroprotective; BAD18ct22; neuroleptic; hypotensive; gene therapy; BAD18cag1; hypertension; neuropsychiatric disorder; bipolar affective disorder; schizophrenia; bipolar mood disorder; manic-depressive illness; Alzheimer's disease; neurodegenerative disorder; senile dementia; Huntington's disease; amyotrophic lateral sclerosis; Parkinson's disease; sleep disorder; Gilles de la Tourette's syndrome; ds.

XX OS Homo sapiens.

XX FH Key

FT misc_feature

FT Location/Qualifiers

FT 28441..29265

FT /*tag= a

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 29683..39587

FT /*tag= b

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 40284..43253

FT /*tag= c

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 43518..46075

FT /*tag= d

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 47264..52284

FT /*tag= e

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 52672..56935

FT /*tag= f

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 57032..57726

FT /*tag= g

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 58065..59057

FT /*tag= h

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 59815..60471

FT /*tag= i

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 60870..62451

FT /*tag= j

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 62543..63268

FT /*tag= k

FT /*note= "This sequence is specifically claimed in Claim 1"

FT 63494..66959

FT /*tag= l

FT /*note= "This sequence is specifically claimed in Claim 1"


```
XX WPI; 2001-656860/75.
DR P-PSDB; ABB65081.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 22034; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4924 BP; 1405 A; 1046 C; 1111 G; 1362 T; 0 other;
SQ

Query Match 68.0%; Score 20.4; DB 23; Length 4924;
Best Local Similarity 80.0%; Pred. No. 29;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cctgatgcaaatgcccaattgcaggtaa 30
   ||| || ||||| ||||| ||||| |||
Db 4777 CCTTATAGCAAAATACCCAATTGCAATTAA 4748

RESULT 11
ABL20106/c
ID ABL20106 standard; DNA; 4936 BP.
XX
XX ABL20106;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11791.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 11791; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4924 BP; 1405 A; 1046 C; 1111 G; 1362 T; 0 other;
SQ

Query Match 68.0%; Score 20.4; DB 23; Length 4924;
Best Local Similarity 80.0%; Pred. No. 29;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cctgatgcaaatgcccaattgcaggtaa 30
   ||| || ||||| ||||| ||||| |||
Db 4777 CCTTATAGCAAAATACCCAATTGCAATTAA 4748

RESULT 12
ABL10884/c
ID ABL10884 standard; cDNA; 2576 BP.
XX
XX ABL10884;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27134.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 27134; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2576 BP; 547 A; 678 C; 652 G; 699 T; 0 other;
SQ
```

```

Query Match          66.7%; Score 20; DB 23; Length 2576;
Best Local Similarity 82.1%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 cctgatggcaaatgcccaattgcaggt 28
        ||| ||||| ||||| | ||||| | |
Db       393 CCGATGGCAAAATGCTCGAATTGCTGTT 366

RESULT 13
ABL14969
ID ABL14969 standard; cDNA; 3219 BP.
XX
AC ABL14969;
XX
DT DT
XX
DE DE
XX
KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 39389.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PP 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PP WPI; 2001-656860/75.
XX
DR P-PSDB; ABB70866.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 39389; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3219 BP; 810 A; 827 C; 881 G; 701 T; 0 other;

Query Match          66.7%; Score 20; DB 23; Length 3219;
Best Local Similarity 82.1%; Pred. No. 41;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 cctgatggcaaatgcccaattgcaggt 28
        ||| ||||| ||||| | ||||| | |
Db       395 ccgcatggcaaatgctcgaattgctgtt 422

RESULT 14
ABL14968/c
ID ABL14968 standard; cDNA; 5679 BP.

```

OS Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH 841..900
FT misc_feature /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
XX
PN EP786519-A2.
XX
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX
XX Claim 1; Page 756-757; 327lpp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
XX Sequence 2375 BP; 826 A; 336 C; 353 G; 793 T; 67 other;

Query Match 63.3%; Score 19; DB 18; Length 2375;
Best Local Similarity 81.5%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 tgatgcaaatgcccaattgcaggta 29
||||| ||||| || ||||| |||||
Db 2059 tgatgcaaaagcaccattgaagata 2085

Search completed: August 24, 2002, 22:11:51
Job time: 11961 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:55:39 ; Search time 204.64 Seconds
(without alignments)
36.010 Million cell updates/sec

Title: US-09-986-381-2

Perfect score: 30

Sequence: 1 cctgatggcaaatgcccaattgcaggtaa 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/1/ina/PCRUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	17.8	59.3	975	1	US-08-746-797-1 Sequence 1, Appl
2	17.8	59.3	975	1	US-08-927-387-1 Sequence 1, Appl
3	17.8	59.3	975	2	US-08-918-058-1 Sequence 1, Appl
4	17.6	58.7	56	1	US-08-461-184-10 Sequence 10, Appl
5	17.6	58.7	56	1	US-08-463-675-10 Sequence 10, Appl
6	17.6	58.7	56	1	US-08-464-589-10 Sequence 10, Appl
7	17.6	58.7	85	1	US-08-425-336-131 Sequence 131, App
8	17.6	58.7	85	1	US-08-488-113B-131 Sequence 131, App
9	17.6	58.7	85	1	US-08-488-113B-144 Sequence 144, App
10	17.6	58.7	85	1	US-08-477-484B-131 Sequence 131, App
11	17.6	58.7	85	1	US-08-477-484B-144 Sequence 144, App
12	17.6	58.7	85	1	US-08-107-669D-36 Sequence 36, Appl
13	17.6	58.7	85	1	US-08-107-669D-56 Sequence 56, Appl
14	17.6	58.7	85	1	US-08-472-788A-36 Sequence 36, Appl
15	17.6	58.7	85	1	US-08-472-788A-56 Sequence 56, Appl
16	17.6	58.7	85	1	US-08-477-531B-36 Sequence 36, Appl
17	17.6	58.7	85	1	US-08-477-531B-56 Sequence 56, Appl
18	17.6	58.7	85	2	US-08-646-360-131 Sequence 131, App
19	17.6	58.7	85	2	US-08-646-360-173 Sequence 173, App
20	17.6	58.7	85	2	US-08-082-842A-36 Sequence 36, Appl
21	17.6	58.7	85	2	US-08-082-842A-56 Sequence 56, Appl
22	17.6	58.7	85	3	US-08-839-765-131 Sequence 131, App
23	17.6	58.7	85	3	US-08-839-765-144 Sequence 144, App
24	17.6	58.7	85	3	US-09-136-389-131 Sequence 131, App
25	17.6	58.7	85	3	US-09-136-389-173 Sequence 173, App
26	17.6	58.7	425	1	US-08-107-669D-46 Sequence 46, Appl
27	17.6	58.7	425	1	US-08-472-788A-46 Sequence 46, Appl

28	17.6	58.7	425	1	US-08-477-531B-46 Sequence 46, Appl
29	17.6	58.7	425	2	US-08-082-842A-46 Sequence 46, Appl
30	17.6	58.7	435	4	US-08-569-147-75 Sequence 75, Appl
31	17.4	58.0	2610	1	US-08-374-834-17 Sequence 17, Appl
32	17.4	58.0	2610	1	US-08-644-271-28 Sequence 28, Appl
33	17	56.7	20303	1	US-08-370-975B-6 Sequence 6, Appl
34	17	56.7	26764	1	US-08-370-975B-1 Sequence 1, Appl
35	17	56.7	87350	3	US-08-781-891-79 Sequence 79, Appl
36	16.8	56.0	1416	3	US-08-866-928B-2 Sequence 2, Appl
37	16.8	56.0	3249	1	US-08-106-493A-1 Sequence 1, Appl
38	16.8	56.0	3249	1	US-08-429-264-1 Sequence 1, Appl
39	16.8	56.0	4853	1	US-08-832-883-1 Sequence 1, Appl
40	16.8	56.0	4853	2	US-08-832-877-1 Sequence 1, Appl
41	16.8	56.0	4875	1	US-08-460-739-1 Sequence 1, Appl
42	16.6	55.3	786	4	US-08-635-928-31 Sequence 31, Appl
43	16.6	55.3	865	1	US-08-341-568-4 Sequence 4, Appl
44	16.6	55.3	865	2	US-08-911-020-4 Sequence 4, Appl
45	16.6	55.3	1425	1	US-08-464-148-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-746-797-1
; Sequence 1, Application US/08746797
; Patent No. 5759832
; GENERAL INFORMATION:
; APPLICANT: Gentry, Daniel
; APPLICANT: Lonsdale, John
; APPLICANT: Payne, David
; APPLICANT: Pearson, Stewart
; TITLE OF INVENTION: NOVEL FabH
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: U.S.A.
; Zip: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,797
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5759832 Yet Assigned
; FILING DATE: 23-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmli, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50573
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-746-797-1

Query Match	59.3%	Score 17.8	DB 1	Length 975
Best Local Similarity	75.9%	Pred. No. 28		
Matches	22	Conservative 0	Mismatches 7	Indels 0
Gaps	0			
Qy	2	ctgatggcaaatgcccaattgcaggtaa 30		
Db	184	CTGATGCGAAAGCTGGAATAACAGGAAA 212		
RESULT 2				
US-08-927-387-1				
Sequence 1, Application US/08927387				
Patent No. 5783432				
GENERAL INFORMATION:				
APPLICANT: Gentry, Daniel				
APPLICANT: Lonsdale, John				
APPLICANT: Payne, David				
APPLICANT: Pearson, Stewart				
TITLE OF INVENTION: NOVEL FABH				
NUMBER OF SEQUENCES: 7				
CORRESPONDENCE ADDRESS:				
ADDRESSEE: SmithKline Beecham Corporation				
STREET: 709 Swedeland Road				
CITY: King of Prussia				
STATE: PA				
COUNTRY: U.S.A.				
ZIP: 19406-0939				
COMPUTER READABLE FORM:				
MEDIUM TYPE: Diskette				
COMPUTER: IBM Compatible				
OPERATING SYSTEM: DOS				
SOFTWARE: FastSEQ Version 1.5				
CURRENT APPLICATION DATA:				
APPLICATION NUMBER: US/08/927,387				
FILING DATE: 25-AUG-1997				
CLASSIFICATION: 435				
PRIOR APPLICATION DATA:				
APPLICATION NUMBER: 08/746,797				
FILING DATE:				
ATTORNEY/AGENT INFORMATION:				
NAME: Gimmi, Edward R				
REGISTRATION NUMBER: 38,891				
REFERENCE/DOCKET NUMBER: P50573				
TELECOMMUNICATION INFORMATION:				
TELEPHONE: 610-270-4478				
TELEFAX: 610-270-5090				
TELEX:				
INFORMATION FOR SEQ ID NO: 1:				
SEQUENCE CHARACTERISTICS:				
LENGTH: 975 base pairs				
TYPE: nucleic acid				
STRANDEDNESS: double				
TOPOLOGY: linear				
MOLECULE TYPE: Genomic DNA				
HYPOTHETICAL: NO				
ANTI-SENSE: NO				
FRAGMENT TYPE:				
ORIGINAL SOURCE:				
US-08-927-387-1				
Query Match	59.3%	Score 17.8	DB 1	Length 975
Best Local Similarity	75.9%	Pred. No. 28		
Matches	22	Conservative 0	Mismatches 7	Indels 0
Gaps	0			
Qy	2	ctgatggcaaatgcccaattgcaggtaa 30		
Db	184	CTGATGCGAAAGCTGGAATAACAGGAAA 212		
RESULT 3				
US-08-927-387-1				
Sequence 1, Application US/08927387				
Patent No. 5783432				
GENERAL INFORMATION:				
APPLICANT: Gentry, Daniel				
APPLICANT: Lonsdale, John				
APPLICANT: Payne, David				
APPLICANT: Pearson, Stewart				
TITLE OF INVENTION: NOVEL FABH				
NUMBER OF SEQUENCES: 7				
CORRESPONDENCE ADDRESS:				
ADDRESSEE: SmithKline Beecham Corporation				
STREET: 709 Swedeland Road				
CITY: King of Prussia				
STATE: PA				
COUNTRY: U.S.A.				
ZIP: 19406-0939				
COMPUTER READABLE FORM:				
MEDIUM TYPE: Diskette				
COMPUTER: IBM Compatible				
OPERATING SYSTEM: DOS				
SOFTWARE: FastSEQ Version 1.5				
CURRENT APPLICATION DATA:				
APPLICATION NUMBER: US/08/927,387				
FILING DATE: 25-AUG-1997				
CLASSIFICATION: 435				
PRIOR APPLICATION DATA:				
APPLICATION NUMBER: 08/746,797				
FILING DATE: 23-OCT-1996				
ATTORNEY/AGENT INFORMATION:				
NAME: Gimmi, Edward R				
REGISTRATION NUMBER: 38,891				
REFERENCE/DOCKET NUMBER: P50573				
TELECOMMUNICATION INFORMATION:				

```
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 07148
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,184
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,498
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: KELLEY, ROBIN D
; REGISTRATION NUMBER: 34,637
; REFERENCE/DOCKET NUMBER: CRP093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..56
; OTHER INFORMATION: /note= "520C9SIG"
; US-08-461-184-10

Query Match 58.7%; Score 17.6; DB 1; Length 56;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgc 24
||||| ||||| ||||| |||||
DB 27 CCTGATGGCAGTGCCTCCAAAGTGC 50

RESULT 5
US-08-463-675-10
; Sequence 10, Application US/08463675
; Patent No. 5658763
; GENERAL INFORMATION:
; APPLICANT: DORAI, HAIMANTI
; APPLICANT: OPPERMANN, HERMANN
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR HIGH PROTEIN
; TITLE OF INVENTION: PRODUCTION FROM NON-NATIVE DNA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES, INC
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 07148
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,675
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/143,498
; FILING DATE: 25-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KELLEY, ROBIN D
; REGISTRATION NUMBER: 34,637
; REFERENCE/DOCKET NUMBER: CRP093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..56
; OTHER INFORMATION: /note= "520C9SIG"
; US-08-463-675-10

Query Match 58.7%; Score 17.6; DB 1; Length 56;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgc 24
||||| ||||| ||||| |||||
DB 27 CCTGATGGCAGTGCCTCCAAAGTGC 50

RESULT 6
US-08-464-589-10
; Sequence 10, Application US/08464589
; Patent No. 5733782
; GENERAL INFORMATION:
; APPLICANT: DORAI, HAIMANTI
; APPLICANT: OPPERMANN, HERMANN
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR HIGH PROTEIN
; TITLE OF INVENTION: PRODUCTION FROM NON-NATIVE DNA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES, INC
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 07148
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,589
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/143,498
; FILING DATE: 25-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KELLEY, ROBIN D
; REGISTRATION NUMBER: 34,637
; REFERENCE/DOCKET NUMBER: CRP093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..56
; OTHER INFORMATION: /note= "520C9SIG"
US-08-464-589-10

Query Match      58.7%; Score 17.6; DB 1; Length 56;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgc 24
Db 27 CCTGATGGCAGCTGCCCAAAAGTGC 50

RESULT 7
US-08-425-336-131
; Sequence 131, Application US/08425336
; Patent No. 5621083
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,336
; FILING DATE: 18-APR-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,691
; FILING DATE: 12-MAY-1993
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Thomas C.
; REGISTRATION NUMBER: P-36,989
; REFERENCE/DOCKET NUMBER: 31394
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-425-336-131

Query Match      58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgc 24
Db 27 CCTGATGGCAGCTGCCCAAAAGTGC 60

RESULT 8
US-08-488-113B-131
; Sequence 131, Application US/08488113B
; Patent No. 5744580
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,113B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/707-8889
; TELEFAX: 312/707-9155
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-488-113B-131

Query Match      58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgc 24
Db 27 CCTGATGGCAGCTGCCCAAAAGTGC 60
```

Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 9

US-08-488-113B-144
; Sequence 144, Application US/08488113B
; Patent No. 5744580
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,113B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/707-8889
; TELEFAX: 312/707-9155
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 144:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-488-113B-144

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgc 24

Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 10

US-08-477-484B-131
; Sequence 131, Application US/08477484B
; Patent No. 5756699
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnika, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,484B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/707-8889
; TELEFAX: 312/707-9155
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-477-484B-131

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgc 24

Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 11

US-08-477-484B-144
; Sequence 144, Application US/08477484B
; Patent No. 5756699
; GENERAL INFORMATION:

APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 169
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,484B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/425,336
FILING DATE: 18-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 144:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-477-484B-144

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgc 24
||||| ||||| ||||| |||||
DB 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 12
US-08-107-669D-36
Sequence 36, Application US/08107669D
Patent No. 5766886
GENERAL INFORMATION:
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,669D
FILING DATE: 13-AUG-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906

ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,669D
FILING DATE: 13-AUG-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,464
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Michele A. Cimbala
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0610.1000001/MAC
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-107-669D-36

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccccaattgc 24
||||| ||||| ||||| |||||
DB 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 13
US-08-107-669D-56
Sequence 56, Application US/08107669D
Patent No. 5766886
GENERAL INFORMATION:
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,669D
FILING DATE: 13-AUG-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906

```
/ FILING DATE: 14-DEC-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/808,464
/ FILING DATE: 13-DEC-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Michele A. Cimbala
/ REGISTRATION NUMBER: 33,851
/ REFERENCE/DOCKET NUMBER: 0610.1000001/MAC
/ TELEPHONE: 202/371-2600
/ TELEFAX: 202/371-2540
/ INFORMATION FOR SEQ ID NO: 56:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 85 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
/ US-08-107-669D-56

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccccaattgc 24
Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 14
US-08-472-788A-36
; Sequence 36, Application US/08472788A
; Patent No. 5770196
; GENERAL INFORMATION:
; APPLICANT: Studnicka, Gary M.
; TITLE OF INVENTION: Modified Antibody Variable Domains
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,788A
; FILING DATE: 14-DEC-1992
; APPLICATION NUMBER: PCT/US92/10906
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/808,464
; FILING DATE: 13-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0610.1000003
; TELEPHONE: 202/371-2600
; TELEFAX: 202/371-2540
; TELEX:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-472-788A-36
```

```
/ LENGTH: 85 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
/ US-08-472-788A-36

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccccaattgc 24
Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

RESULT 15
US-08-472-788A-56
; Sequence 56, Application US/08472788A
; Patent No. 5770196
; GENERAL INFORMATION:
; APPLICANT: Studnicka, Gary M.
; TITLE OF INVENTION: Modified Antibody Variable Domains
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,788A
; FILING DATE: 14-DEC-1992
; APPLICATION NUMBER: PCT/US92/10906
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/808,464
; FILING DATE: 13-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0610.1000003
; TELEPHONE: 202/371-2600
; TELEFAX: 202/371-2540
; TELEX:
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-472-788A-56

Query Match 58.7%; Score 17.6; DB 1; Length 85;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccccaattgc 24
```

Db 37 CCTGATGGCAGCTGCCCAAGTGC 60

Search completed: August 24, 2002, 21:55:40
Job time: 13135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:08:02 ; Search time 7654.23 Seconds
(without alignments)
52.900 Million cell updates/sec

Title: US-09-986-381-2

Perfect score: 30

Sequence: 1 cctgatggcaaatgcccgaattgcaggtaa 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.4	68.0	486	9	AW430195	AW430195 69344 MAR
C 2	20	66.7	379	9	AA830659	AA830659 OC52h03.S
C 3	20	66.7	435	10	BI209197	BI209197 EST527237
C 4	20	66.7	446	10	BI203400	BI203400 EST521440
C 5	20	66.7	446	10	BI206358	BI206358 EST524398
C 6	20	66.7	446	10	BI210510	BI210510 EST528550
C 7	20	66.7	547	10	BI196241	BI196241 BJ196241
C 8	20	66.7	550	9	AI389978	AI389978 GH21414.5
C 9	20	66.7	554	10	BG36328	BG36328 SD13967.5
C 10	20	66.7	720	10	BG328081	BG328081 602427131
C 11	20	66.7	772	9	AI107550	AI107550 GH05289.5
C 12	20	66.7	1045	12	CNS0182H	ALI08803 Drosophila
C 13	19.6	65.3	169	10	BJ187414	BJ187414 BJ187414
C 14	19.6	65.3	176	10	BJ195730	BJ195730 BJ195730
C 15	19.6	65.3	280	12	AQ069046	AQ069046 HS_2255.B
C 16	19.6	65.3	335	10	BJ196713	BJ196713 BJ196713
C 17	19.6	65.3	363	10	BJ176443	BJ176443 BJ176443

C 18	19.6	65.3	368	10	BJ202332	BJ202332
C 19	19.6	65.3	382	10	BJ159950	BJ159950
C 20	19.6	65.3	407	10	BM278380	BM278380 As_tgz_58
C 21	19.6	65.3	425	10	BJ194714	BJ194714
C 22	19.6	65.3	434	10	BJ206663	BJ206663
C 23	19.6	65.3	440	9	AT006512	AT006512
C 24	19.6	65.3	440	10	BJ185283	BJ185283
C 25	19.6	65.3	452	10	BJ185987	BJ185987
C 26	19.6	65.3	455	10	BJ185441	BJ185441
C 27	19.6	65.3	456	10	BJ162443	BJ162443
C 28	19.6	65.3	462	10	BJ162586	BJ162586
C 29	19.6	65.3	462	10	BJ181571	BJ181571
C 30	19.6	65.3	463	10	BJ185841	BJ185841
C 31	19.6	65.3	464	10	BJ174484	BJ174484
C 32	19.6	65.3	464	10	BJ194801	BJ194801
C 33	19.6	65.3	465	10	BJ161237	BJ161237
C 34	19.6	65.3	468	10	BJ162435	BJ162435
C 35	19.6	65.3	470	10	BJ186557	BJ186557
C 36	19.6	65.3	474	10	BJ158198	BJ158198
C 37	19.6	65.3	475	10	BJ159429	BJ159429
C 38	19.6	65.3	475	10	BJ160020	BJ160020
C 39	19.6	65.3	476	10	BJ199030	BJ199030
C 40	19.6	65.3	485	10	BJ198104	BJ198104
C 41	19.6	65.3	489	10	BJ159509	BJ159509
C 42	19.6	65.3	490	10	BJ162370	BJ162370
C 43	19.6	65.3	495	10	BJ201664	BJ201664
C 44	19.6	65.3	496	10	BJ162712	BJ162712
C 45	19.6	65.3	496	10	BJ197147	BJ197147

ALIGNMENTS

RESULT 1
AW430195/c
LOCUS 69344 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION AW430195
ACCESSION AW430195
VERSION AW430195.1 GI:6961502
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 486)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Cassas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 45 row: K column: 1
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
Source
1..486
/organism="Bos taurus"
/db_xref="taxon:9913"

```

/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: PCMW SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      173 a      69 c      77 g      167 t
ORIGIN

Query Match      68.0%; Score 20.4; DB 9; Length 486;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 cctgagcgcaaatgcccaattgcagtaa 30
||||| ||||| ||||| ||||| |||||
Db 480 CCTGATGACATATGCTACAAATTCATGTAA 451

RESULT 2
AA830659      379 bp      mRNA      linear      EST 18-MAR-1998
LOCUS
DEFINITION    ocs2h03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1353365 3',
              mRNA sequence.
ACCESSION     AA830659
VERSION       AA830659
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 379)
AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
              Ph.D., Gerald Marti, M.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bonaldo, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert length: 1117 Std Error: 0.00
              Seq primer: -40ml3 fwd. ET from Amersham
              High quality sequence stop: 292.
              Location/Qualifiers
              1..379
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:1353365"
              /clone_lib="NCI_CGAP_GCB1"
              /tissue_type="germinal center B cell"
              /lab_host="DH10B"
              /note="vector: pT7T3D-Pac (Pharmacia) with a modified
              polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
              was prepared from human tonsillar cells enriched for
              germinal center B cells by flow sorting (CD20+, IgD-),
              provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
              (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
              primed with a Not I - oligo(dT) primer
              [5'-TGTTACCAATCTCAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
              ]. Double-stranded cDNA was ligated to Eco RI adaptors
              (Pharmacia), digested with Not I and cloned into the Not I
              and Eco RI sites of the modified pT7T3 vector. Library
              went through one round of normalization, and was
              constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      131 a      56 c      46 g      146 t
ORIGIN

```

```

Query Match      66.7%; Score 20; DB 9; Length 379;
Best Local Similarity 82.1%; Pred. No. 3.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 tgatggcaaatgcccaattgcagtaa 30
||||| ||||| ||||| ||||| |||||
Db 267 TGATTGCAAAATCCCAACTGTAGTAA 294

RESULT 3
BI209197      435 bp      mRNA      linear      EST 11-JUL-2001
LOCUS
DEFINITION    EST527237 cTOS Lycopersicon esculentum cDNA clone cTOS19018 5' end,
              mRNA sequence.
ACCESSION     BI209197
VERSION       BI209197.1 GI:14686921
KEYWORDS      EST.
SOURCE        tomato.
ORGANISM      Lycopersicon esculentum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
              Lycopersicon.
REFERENCE     1 (bases 1 to 435)
AUTHORS       van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
              Ronning,C. and Tanksley,S.
              Generation of ESTs from Tomato Suspension Cultures
              Unpublished (2001)
JOURNAL       Contact: CUGI
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Email: http://www.genome.clemson.edu/orders/index.html.
              Location/Qualifiers
              1..435
              /organism="Lycopersicon esculentum"
              /cultivar="TA496, E6203"
              /db_xref="taxon:4081"
              /clone="cTOS19018"
              /clone_lib="cTOS"
              /tissue_type="suspension cultures"
              /lab_host="SOLR"
              /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
              XhoI; Suspension cultures of L.esculentum E6203 were grown
              in Murashige and Skoog based medium, supplemented with 1%
              coconut milk (filter sterilized and added after
              autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
              Fresh medium was added every 7 days, and cultures were
              grown at 25 C, with 12hrs of light and continuous
              shaking."
BASE COUNT      129 a      89 g      123 t
ORIGIN

Query Match      66.7%; Score 20; DB 10; Length 435;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ctgatggcaaatgcccaattgcagta 29
||||| ||||| ||||| ||||| ||
Db 290 CAGATGGCAAAATCCCTGAATGCAGATA 263

RESULT 4
BI203400/c
LOCUS
DEFINITION    EST521440 cTOS Lycopersicon esculentum cDNA clone cTOS1L21 5' end,
              mRNA sequence.
ACCESSION     BI203400
VERSION       BI203400.1 GI:14681124
KEYWORDS      EST.

```

```

SOURCE
ORGANISM      Lycopersicon esculentum
               tomato.
REFERENCE
AUTHORS       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
               Lycopersicon.
TITLE         1 (bases 1 to 446)
JOURNAL       van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
COMMENT       Ronning,C. and Tanksley,S.
               Generation of ESTs from Tomato Suspension Cultures
               Unpublished (2001)
               Contact: CUGI
               Clemson University
               Clemson University
               100 Jordan Hall, Clemson, SC 29634, USA
               Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
1. .446
   /organism="Lycopersicon esculentum"
   /cultivar="TA496, E6203"
   /db_xref="taxon:4081"
   /clone="cTOS1L21"
   /tissue_type="suspension cultures"
   /lab_host="SOLR"
   /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
   XhoI; Suspension cultures of L.esculentum E6203 were grown
   in Murashige and Skoog based medium, supplemented with 15%
   coconut milk (filter sterilized and added after
   autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
   Fresh medium was added every 7 days, and cultures were
   grown at 25 C, with 12hrs of light and continuous
   shaking."
BASE COUNT    134 a 96 c 91 g 125 t
ORIGIN

Query Match   66.7%; Score 20; DB 10; Length 446;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  2 ctgatggcaaatgcccattgcaggta 29
      | ||||| ||||| || ||||| ||
DB  301 CAGATGGCAATGCCTGAATGCAGATA 274

RESULT  5
BI206358/c
LOCUS      446 bp mRNA linear EST 11-JUL-2001
DEFINITION EST524398 cTOS Lycopersicon esculentum cDNA clone cTOS10B12 5' end,
           mRNA sequence.
ACCESSION  BI206358
VERSION    BI206358.1 GI:14684082
KEYWORDS   tomato.
SOURCE     Lycopersicon esculentum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
           Lycopersicon.
REFERENCE  1 (bases 1 to 446)
AUTHORS   van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
           Ronning,C. and Tanksley,S.
           Generation of ESTs from Tomato Suspension Cultures
           Unpublished (2001)
TITLE     Contact: CUGI
JOURNAL   Clemson University
COMMENT   Clemson University
           100 Jordan Hall, Clemson, SC 29634, USA
           Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
1. .446
   /organism="Lycopersicon esculentum"
   /cultivar="TA496, E6203"
   /db_xref="taxon:4081"
   /clone="cTOS1L21"
   /tissue_type="suspension cultures"
   /lab_host="SOLR"
   /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
   XhoI; Suspension cultures of L.esculentum E6203 were grown
   in Murashige and Skoog based medium, supplemented with 15%
   coconut milk (filter sterilized and added after
   autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
   Fresh medium was added every 7 days, and cultures were
   grown at 25 C, with 12hrs of light and continuous
   shaking."
BASE COUNT    134 a 96 c 91 g 125 t
ORIGIN

Query Match   66.7%; Score 20; DB 10; Length 446;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  2 ctgatggcaaatgcccattgcaggta 29
      | ||||| ||||| || ||||| ||
DB  301 CAGATGGCAATGCCTGAATGCAGATA 274

RESULT  5
BI206358/c
LOCUS      446 bp mRNA linear EST 11-JUL-2001
DEFINITION EST524398 cTOS Lycopersicon esculentum cDNA clone cTOS10B12 5' end,
           mRNA sequence.
ACCESSION  BI206358
VERSION    BI206358.1 GI:14684082
KEYWORDS   tomato.
SOURCE     Lycopersicon esculentum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
           Lycopersicon.
REFERENCE  1 (bases 1 to 446)
AUTHORS   van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
           Ronning,C. and Tanksley,S.
           Generation of ESTs from Tomato Suspension Cultures
           Unpublished (2001)
TITLE     Contact: CUGI
JOURNAL   Clemson University
COMMENT   Clemson University
           100 Jordan Hall, Clemson, SC 29634, USA
           Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
1. .446
   /organism="Lycopersicon esculentum"

```

```

/cultivar="TA496, E6203"
/db_xref="taxon:4081"
/clone="cTOS10B12"
/clone_lib="cTOS"
/tissue_type="suspension cultures"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."
BASE COUNT    134 a 96 c 91 g 125 t
ORIGIN

Query Match   66.7%; Score 20; DB 10; Length 446;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  2 ctgatggcaaatgcccattgcaggta 29
      | ||||| ||||| || ||||| ||
DB  301 CAGATGGCAATGCCTGAATGCAGATA 274

RESULT  6
BI210510/c
LOCUS      446 bp mRNA linear EST 11-JUL-2001
DEFINITION EST528550 cTOS Lycopersicon esculentum cDNA clone cTOS23B23 5' end,
           mRNA sequence.
ACCESSION  BI210510
VERSION    BI210510.1 GI:14688234
KEYWORDS   tomato.
SOURCE     Lycopersicon esculentum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
           Lycopersicon.
REFERENCE  1 (bases 1 to 446)
AUTHORS   van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
           Ronning,C. and Tanksley,S.
           Generation of ESTs from Tomato Suspension Cultures
           Unpublished (2001)
TITLE     Contact: CUGI
JOURNAL   Clemson University
COMMENT   Clemson University
           100 Jordan Hall, Clemson, SC 29634, USA
           Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
1. .446
   /organism="Lycopersicon esculentum"
   /cultivar="TA496, E6203"
   /db_xref="taxon:4081"
   /clone="cTOS23B23"
   /tissue_type="suspension cultures"
   /lab_host="SOLR"
   /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
   XhoI; Suspension cultures of L.esculentum E6203 were grown
   in Murashige and Skoog based medium, supplemented with 15%
   coconut milk (filter sterilized and added after
   autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
   Fresh medium was added every 7 days, and cultures were
   grown at 25 C, with 12hrs of light and continuous
   shaking."
BASE COUNT    134 a 96 c 91 g 125 t
ORIGIN

Query Match   66.7%; Score 20; DB 10; Length 446;

```

```

Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ctgatggcaaatgcccaattgcaggta 29
|||||
Db 301 CAGATGGCAAATGCTGAAATGCAGATA 274

RESULT 7
BJ196241/c
LOCUS
DEFINITION
BJ196241 547 bp mRNA linear EST 24-JAN-2002
BJ196241 normalized full length cDNA library, chloronemata,
caulonemata and rhizoid-like protonemata Physcomitrella patens
subsp. patens cDNA clone ppIn26p06 5', mRNA sequence.
ACCESSION
VERSION BJ196241.1 GI:183364168
KEYWORDS
SOURCE
ORGANISM Physcomitrella patens subsp. patens.
Physcomitrella patens subsp. patens.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 547)
REFERENCE
AUTHORS Fujita,T., Shin-i.T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
,Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
,M.
Comparison of the moss Physcomitrella patens genome with flowering
plants genome
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genetics.nig.ac.jp
A backbone of the vector is pBluescript II, that was in vivo
excised from a modified LPS phage vector (Mo bi Tec, Germany). XhoI
digested-5' end of cDNA is ligated to SalI site of the vector, and
the BamHI digested-3' end including poly-A tail is ligated to BamHI
site of the vector. cDNA insert could be amplified with
conventional T7 and T3 primers. This normalized full-length cDNA
library was generated basically according to the method described
in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
Protonemata were blended by the POLYTRON, and then cultivated on
the BCD medium containing lum NAA (naphthalene acetic acid) for 8
to 11 days under the continuous light.
FEATURES
source
1..547
Location/Qualifiers
/organism="Physcomitrella patens subsp. patens"
/db_xref="taxon:145481"
/clone_lib="ppIn26p06"
/tissue_type="mixture of chloronemata, caulonemata and
rhizoid-like protonemata"
BASE COUNT 137 a 138 c 152 g 120 t
ORIGIN

Query Match 66.7%; Score 20; DB 10; Length 547;
Best Local Similarity 82.1%; Pred. No. 3.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ctgatggcaaatgcccaattgcaggta 29
|||||
Db 160 CTGCTGCCAAATGCTCAACTGCAGGGA 133

RESULT 8
AI389978
LOCUS
DEFINITION
GH21414.5prime GH Drosophila melanogaster head pot2
melanogaster cDNA clone GH21414 5prime, mRNA sequence.

AI389978 550 bp mRNA linear EST 19-APR-2001
GH21414.5prime GH Drosophila melanogaster head pot2 Drosophila
melanogaster cDNA clone GH21414 5prime, mRNA sequence.

```

```

AI389978
LOCUS
DEFINITION
SDI3967.5prime SD Drosophila melanogaster Schneider I2 cell culture
pot2 Drosophila melanogaster cDNA clone SDI3967 5 similar to
CG9201: FBan0009201 located on: X 13D2-13D2.; 04/13/2001, mRNA
sequence.
AI389978 554 bp mRNA linear EST 23-APR-2001
SDI3967.5prime SD Drosophila melanogaster Schneider I2 cell culture
pot2 Drosophila melanogaster cDNA clone SDI3967 5 similar to
CG9201: FBan0009201 located on: X 13D2-13D2.; 04/13/2001, mRNA
sequence.
BASE COUNT 160 a 134 c 160 g 96 t
ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 550;
Best Local Similarity 82.1%; Pred. No. 3.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcagggt 28
|||||
Db 394 CCGGATGGCAAATGCTCGAATGCTGTT 421

RESULT 9
BG636328
LOCUS
DEFINITION
BG636328 554 bp mRNA linear EST 23-APR-2001
SDI3967.5prime SD Drosophila melanogaster Schneider I2 cell culture
pot2 Drosophila melanogaster cDNA clone SDI3967 5 similar to
CG9201: FBan0009201 located on: X 13D2-13D2.; 04/13/2001, mRNA
sequence.
ACCESSION
VERSION BG636328.1 GI:13763865
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 554)
REFERENCE
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
High quality sequence stop: 514.
FEATURES
source
1..550
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT 160 a 134 c 160 g 96 t
ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 550;
Best Local Similarity 82.1%; Pred. No. 3.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcagggt 28
|||||
Db 394 CCGGATGGCAAATGCTCGAATGCTGTT 421

RESULT 9
BG636328
LOCUS
DEFINITION
BG636328 554 bp mRNA linear EST 23-APR-2001
SDI3967.5prime SD Drosophila melanogaster Schneider I2 cell culture
pot2 Drosophila melanogaster cDNA clone SDI3967 5 similar to
CG9201: FBan0009201 located on: X 13D2-13D2.; 04/13/2001, mRNA
sequence.
ACCESSION
VERSION BG636328.1 GI:13763865
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 554)
REFERENCE
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003499: arm:X [15147196,15448978]

```

estimated-cyto:13C5-13E14: 04/13/2001

Plate: SD.139 row: F column: 7

High quality sequence stop: 474.

Location/Qualifiers

FEATURES

source

1. .554 /organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="SD13967"

/clone_lib="SD Drosophila melanogaster Schneider L2 cell

culture pot2"

/lab_host="DH5-alpha"

/note="vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized

fractionated cDNAs were directly ligated into pOT2.

Plasmid cDNA library."

164 a 132 c 163 g 95 t

BASE COUNT

ORIGIN

Query Match 66.7%; Score 20; DB 10; Length 554;
Best Local Similarity 82.1%; Pred. No. 3.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcaggt 28

Db 391 CCGATGCGCAATGCTCGAATTGCTGT 418

RESULT 10

BG328081

LOCUS

DEFINITION 602427131F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546735 5',

mRNA sequence.

ACCESSION BG328081

VERSION BG328081.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 720)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI233 row: m column: 08

High quality sequence stop: 707.

Location/Qualifiers

1. .720

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4546735"

/clone_lib="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

204 a 132 c 179 g 205 t

BASE COUNT

ORIGIN

Query Match 66.7%; Score 20; DB 10; Length 720;
Best Local Similarity 82.1%; Pred. No. 3.9e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ttagtgcaaatgcccaattgcaggtaa 30

Db 251 TGATGCAAAATACCCAGGAGCAGTTAA 278

RESULT 11

AI107550

LOCUS

DEFINITION

AI107550 772 bp mRNA linear EST 23-APR-2001

GH05289.5prime GH Drosophila melanogaster head pOT2 Drosophila

melanogaster cDNA clone GH05289 5 similar to CG9201: FBan0009201

located on: X 13D2-13D2:: 04/10/2001, mRNA sequence.

AI107550

AI107550.2 GI:13759513

EST.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 772)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

On Aug 26, 1998 this sequence version replaced gi:3475203.

Other_ESTS: GH05289.3prime

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AB003499; arm: X [15147196,15448978]

estimated-cyto:13C5-13E14: 04/10/2001

Plate: GH.52 row: H column: 5

High quality sequence stop: 688

POLYA=No.

Location/Qualifiers

1. .772

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="GH05289"

/clone_lib="GH Drosophila melanogaster head pOT2"

/sex="male and female"

/dev_stage="adult"

/lab_host="DH5 - alpha"

/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2. Plasmid cDNA library."

221 a 189 c 212 g 150 t

BASE COUNT

ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 772;
Best Local Similarity 82.1%; Pred. No. 3.9e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cctgatggcaaatgcccaattgcaggt 28

Db 395 CCGATGCGCAAAATGCTCGAATTGCTGT 422

RESULT 12

CNS0182H

LOCUS

DEFINITION

CNS0182H 1045 bp DNA linear GSS 26-JUL-1999

Drosophila melanogaster genome survey sequence SP6 end of BAC

BACN37001 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL108803
 VERSION AL108803.1 GI:5629107
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1045)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.
 FEATURES
 source Location/Qualifiers
 1..1045
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN37001"
 /note="end : SP6"
 BASE COUNT 291 a 222 c 259 g 190 t 83 others
 ORIGIN
 Query Match 66.7%; Score 20; DB 12; Length 1045;
 Best Local Similarity 82.1%; Pred. No. 4.2e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 cctgatggcaaatgcccaattgcaggt 28
 ||| ||||| ||||| ||||| ||
 Db 451 CCGATGGCAAAATGCTCGAATGCTGTT 478
 RESULT 13
 BJI87414/c
 LOCUS BJI87414 169 bp mRNA linear EST 24-JAN-2002
 DEFINITION BJI87414 normalized full length cDNA library, chloronemata,
 caulonemata and malformed buds Physcomitrella patens subsp. patens
 cDNA clone pphb40a20 5', mRNA sequence.
 ACCESSION BJI87414.1 GI:18355355
 VERSION BJI87414
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens.
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 169)
 AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
 Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
 ,M.
 TITLE Comparison of the moss Physcomitrella patens genome with flowering
 plants genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is pBluescript II, that was in vivo
 excised from a modified IPS phase vector (Mo bi Tec, Germany). XhoI
 digested-5' end of cDNA is ligated to SalI site of the vector, and
 the BamHI digested-3' end including poly-A tail is ligated to BamHI
 site of the vector. cDNA instert could be amplified with
 conventional T7 and T3 primers. This normalized full- length cDNA
 library was generated basically according to the method described
 in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
 protonemata were blended by the POLYTRON, and then cultivated on
 the BCD medium containing luM NAA (naphthalene acetic acid) for 8
 to 11 days under the continuous light.
 FEATURES
 source Location/Qualifiers
 1..176
 /organism="Physcomitrella patens subsp. patens"
 /db_xref="taxon:145481"
 /clone="pphb40a20"
 /clone_lib="normalized full length cDNA library,
 chloronemata, caulonemata and malformed buds"
 /tissue_type="mixture of chloronemata, caulonemata and
 malformed buds"
 BASE COUNT 36 a 35 c 51 g 47 t
 ORIGIN
 Query Match 65.3%; Score 19.6; DB 10; Length 169;
 Best Local Similarity 84.6%; Pred. No. 4.2e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 ctgatggcaaatgcccaattgcaggt 27
 ||| || ||||| ||||| ||||| ||
 Db 156 CTGCTGCCAAATGCTCCAACTGCAGG 131
 RESULT 14
 BJI95730/c
 LOCUS BJI95730 176 bp mRNA linear EST 24-JAN-2002
 DEFINITION BJI95730 normalized full length cDNA library, chloronemata,
 caulonemata and rhizoid-like protonemata Physcomitrella patens
 subsp. patens cDNA clone pphn24k18 5', mRNA sequence.
 ACCESSION BJI95730
 VERSION BJI95730.1 GI:18363658
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens.
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 176)
 AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
 Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
 ,M.
 TITLE Comparison of the moss Physcomitrella patens genome with flowering
 plants genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is pBluescript II, that was in vivo
 excised from a modified IPS phase vector (Mo bi Tec, Germany). XhoI
 digested-5' end of cDNA is ligated to SalI site of the vector, and
 the BamHI digested-3' end including poly-A tail is ligated to BamHI
 site of the vector. cDNA instert could be amplified with
 conventional T7 and T3 primers. This normalized full- length cDNA
 library was generated basically according to the method described
 in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
 protonemata were blended by the POLYTRON, and then cultivated on
 the BCD medium containing luM NAA (naphthalene acetic acid) for 8
 to 11 days under the continuous light.
 FEATURES
 source Location/Qualifiers
 1..176
 /organism="Physcomitrella patens subsp. patens"
 /db_xref="taxon:145481"
 /clone="pphn24k18"

the BamHI digested-3' end including poly-A tail is ligated to BamHI
 site of the vector. cDNA instert could be amplified with
 conventional T7 and T3 primers. This normalized full- length cDNA
 library was generated basically according to the method described
 in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
 Protonemata were blended by the POLYTRON, and then cultivated on
 the BCD medium containing 0.5uM BA (benzylaminopurine) for 8 to 13
 days under the continuous light.
 FEATURES
 source Location/Qualifiers
 1..169
 /organism="Physcomitrella patens subsp. patens"
 /db_xref="taxon:145481"
 /clone="pphb40a20"
 /clone_lib="normalized full length cDNA library,
 chloronemata, caulonemata and malformed buds"
 /tissue_type="mixture of chloronemata, caulonemata and
 malformed buds"

BASE COUNT 36 a 35 c 51 g 47 t
 ORIGIN
 Query Match 65.3%; Score 19.6; DB 10; Length 169;
 Best Local Similarity 84.6%; Pred. No. 4.2e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 ctgatggcaaatgcccaattgcaggt 27
 ||| || ||||| ||||| ||||| ||
 Db 156 CTGCTGCCAAATGCTCCAACTGCAGG 131

RESULT 14
 BJI95730/c
 LOCUS BJI95730 176 bp mRNA linear EST 24-JAN-2002
 DEFINITION BJI95730 normalized full length cDNA library, chloronemata,
 caulonemata and rhizoid-like protonemata Physcomitrella patens
 subsp. patens cDNA clone pphn24k18 5', mRNA sequence.
 ACCESSION BJI95730
 VERSION BJI95730.1 GI:18363658
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens.
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 176)
 AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
 Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
 ,M.
 TITLE Comparison of the moss Physcomitrella patens genome with flowering
 plants genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is pBluescript II, that was in vivo
 excised from a modified IPS phase vector (Mo bi Tec, Germany). XhoI
 digested-5' end of cDNA is ligated to SalI site of the vector, and
 the BamHI digested-3' end including poly-A tail is ligated to BamHI
 site of the vector. cDNA instert could be amplified with
 conventional T7 and T3 primers. This normalized full- length cDNA
 library was generated basically according to the method described
 in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
 protonemata were blended by the POLYTRON, and then cultivated on
 the BCD medium containing luM NAA (naphthalene acetic acid) for 8
 to 11 days under the continuous light.
 FEATURES
 source Location/Qualifiers
 1..176
 /organism="Physcomitrella patens subsp. patens"
 /db_xref="taxon:145481"
 /clone="pphn24k18"

BASE COUNT 36 a 35 c 51 g 47 t
 ORIGIN
 Query Match 65.3%; Score 19.6; DB 10; Length 169;
 Best Local Similarity 84.6%; Pred. No. 4.2e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 ctgatggcaaatgcccaattgcaggt 27
 ||| || ||||| ||||| ||||| ||
 Db 156 CTGCTGCCAAATGCTCCAACTGCAGG 131
 RESULT 14
 BJI95730/c
 LOCUS BJI95730 176 bp mRNA linear EST 24-JAN-2002
 DEFINITION BJI95730 normalized full length cDNA library, chloronemata,
 caulonemata and rhizoid-like protonemata Physcomitrella patens
 subsp. patens cDNA clone pphn24k18 5', mRNA sequence.
 ACCESSION BJI95730
 VERSION BJI95730.1 GI:18363658
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens.
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 176)
 AUTHORS Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
 Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
 ,M.
 TITLE Comparison of the moss Physcomitrella patens genome with flowering
 plants genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is pBluescript II, that was in vivo
 excised from a modified IPS phase vector (Mo bi Tec, Germany). XhoI
 digested-5' end of cDNA is ligated to SalI site of the vector, and
 the BamHI digested-3' end including poly-A tail is ligated to BamHI
 site of the vector. cDNA instert could be amplified with
 conventional T7 and T3 primers. This normalized full- length cDNA
 library was generated basically according to the method described
 in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
 protonemata were blended by the POLYTRON, and then cultivated on
 the BCD medium containing luM NAA (naphthalene acetic acid) for 8
 to 11 days under the continuous light.
 FEATURES
 source Location/Qualifiers
 1..176
 /organism="Physcomitrella patens subsp. patens"
 /db_xref="taxon:145481"
 /clone="pphn24k18"

/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and rhizoid-like protonemata"
/tissue_type="mixture of chloronemata, caulonemata and
rhizoid-like protonemata"

BASE COUNT 41 a 35 c 52 g 46 t 2 others
ORIGIN

Query Match 65.3%; Score 19.6; DB 10; Length 176;
Best Local Similarity 84.6%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ctgatggcaaatgcccaattgcagg 27
||||| ||||| ||||| ||||| |||||
Db 158 CTGCTGCCAAATGCTCCAACTGCAGG 133

RESULT 15
AQ069046 280 bp DNA linear GSS 04-AUG-1998
LOCUS HS_2255_B2_F10_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2255 Col=20 Row=L, DNA sequence.
ACCESSION AQ069046
VERSION AQ069046.1 GI:3384245
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 280)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged Connector
Plate: 2255 row: L column: 20
Class: BAC ends
High quality sequence stop: 280.

FEATURES
Source
1..280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2255 Col=20 Row=L"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 87 a 62 c 66 g 65 t
ORIGIN

Query Match 65.3%; Score 19.6; DB 12; Length 280;
Best Local Similarity 84.6%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cctgatggcaaatgcccaattgcag 26
||||| ||||| ||||| ||||| |||||
Db 77 CATCATGGCAAGGCCCAATTGGAG 102

Search completed: August 24, 2002, 21:08:07
Job time: 17327 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:53:09 ; Search time 2613.74 Seconds
(without alignments)
264.210 Million cell updates/sec

Title: US-09-986-381-3

Perfect score: 33

Sequence: 1 gtcaagtagcatctgtatcaggcaaaagctag-33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl.*

8: gb_pr.*

9: gb_ro.*

10: gb_sy.*

11: gb_un.*

12: gb_vt.*

13: gb_vt.*

14: gb_vt.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sy.*

28: em_un.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgO_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
------------	-----------	-------	-------	--------	----	----	-------------

c	1	33	100.0	133	9	S81486	S81486
c	2	33	100.0	160	9	S66666	S66666
c	3	33	100.0	4360	9	AB018045	AB018045 Homo sapi
c	4	33	100.0	20303	9	X54156	X54156 Human p53 g
c	5	33	100.0	20303	9	U94788	U94788 Human p53 (
c	6	33	100.0	159021	2	AC087388	AC087388 Homo sapi
c	7	33	100.0	160457	2	AC008049	AC008049 Homo sapi
c	8	23	69.7	118985	2	AC087152	AC087152 Mus muscu
c	9	22	66.7	112904	9	AC027342	AC027342 Homo sapi
c	10	22	66.7	159524	9	CNS01DUO	AL133305 Human chr
c	11	22	66.7	184536	2	AC068682	AC068682 Homo sapi
c	12	22	66.7	189563	2	AC106128	AC106128 Homo sapi
c	13	21.8	66.1	98779	2	AC095715	AC095715 Rattus no
c	14	21.8	66.1	145111	2	AL390245	AL390245 Homo sapi
c	15	21.8	66.1	202174	2	AL645507	AL645507 Homo sapi
c	16	21.4	64.8	490	11	G10379	G10379 human STS C
c	17	21.4	64.8	74982	9	AC067958	AC067958 Homo sapi
c	18	21.4	64.8	152533	9	AC008483	AC008483 Homo sapi
c	19	21.4	64.8	152895	2	AC031979	AC031979 Homo sapi
c	20	21.4	64.8	169328	2	AC016281	AC016281 Homo sapi
c	21	21.4	64.8	170817	9	AL513166	AL513166 Human DNA
c	22	21.4	64.8	177106	2	AC068136	AC068136 Homo sapi
c	23	21.4	64.8	183795	9	AC010423	AC010423 Homo sapi
c	24	21	63.6	36521	2	AC020203	AC020203 Drosophil
c	25	21	63.6	67297	3	AC005135	AC005135 Drosophil
c	26	21	63.6	173188	3	AE003196	AE003196 Drosophil
c	27	21	63.6	265236	3	AE003626	AE003626 Drosophil
c	28	20.8	63.0	95983	9	AC004874	AC004874 Homo sapi
c	29	20.8	63.0	100785	2	AC069169	AC069169 Homo sapi
c	30	20.8	63.0	108230	2	AC094267	AC094267 Rattus no
c	31	20.8	63.0	143967	9	AL138744	AL138744 Human DNA
c	32	20.8	63.0	149919	2	AC007953	AC007953 Homo sapi
c	33	20.8	63.0	166407	2	AC103549	AC103549 Homo sapi
c	34	20.8	63.0	185401	2	AC024927	AC024927 Homo sapi
c	35	20.8	63.0	197284	2	AC092897	AC092897 Homo sapi
c	36	20.8	63.0	214622	2	AL646046	AL646046 Mus muscu
c	37	20.8	63.0	245708	9	AF250324	AF250324 Homo sapi
c	38	20.8	62.4	275159	9	UB2670	UB2670 Homo sapien
c	39	20.6	62.4	150780	2	AC106502	AC106502 Rattus no
c	40	20.6	62.4	182975	9	AP003558	AP003558 Homo sapi
c	41	20.6	62.4	188922	2	AC084833	AC084833 Homo sapi
c	42	20.4	61.8	798	8	AF230709	AF230709 Calycanth
c	43	20.4	61.8	1500	9	HUMZFFSREB	M28372 Homo sapien
c	44	20.4	61.8	1549	9	AK054592	AK054592 Homo sapi
c	45	20.4	61.8	1594	9	BC000288	BC000288 Homo sapi

ALIGNMENTS

RESULT	1	S81486	133 bp	DNA	linear	PRI 07-MAY-1993
LOCUS		S81486	p53	{alternatively spliced, intron 9}	[human, Genomic Mutant, 133	
DEFINITION		nt].				
ACCESSION		S81486				
VERSION		S81486.1	GI:245371			
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
		1 (bases 1 to 133)				
		Felix, C.A., Nau, M.M., Takahashi, T., Mitsudomi, T., Chiba, I.,				
		Poplack, D.G., Reaman, G.H., Cole, D.E., Letterio, J.J., Whang-Peng, J.				
		et, al.				
TITLE		Hereditary and acquired p53 gene mutations in childhood acute				
JOURNAL		lymphoblastic leukemia				
MEDLINE		J. Clin. Invest. 89 (2), 640-647 (1992)				
REMARK		GenBank staff at the National Library of Medicine created this				
		entry [NCBI gbbbsq 81486] from the original journal article.				
		This sequence comes from Figure 2.				

COMMENT Map location: chromosome 17 band p13.1.
FEATURES 133 bp insertion intron 9.

source 1. .133
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

gene 1. .133
partial
/gene="p53"

BASE COUNT 45 a 22 c 26 g 40 t
ORIGIN

Query Match 100.0%; Score 33; DB 9; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcaagtagcatctgtatcaggcaagtcataag 33
Db 85 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 53

RESULT 2

LOCUS S66666 160 bp mRNA linear PRI 23-DEC-1993
DEFINITION p53-tumor suppressor {alternatively spliced, exon 9-10} [human,
Molt-4, T-lymphoblastic leukemia cell line, mRNA PartialMutant, 160
nt].

ACCESSION S66666
VERSION S66666.1 GI:436292

KEYWORDS human Molt-4 T-lymphoblastic leukemia cell line.
SOURCE
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Chow.V.T., Quek.H.H. and Tock.E.P.
TITLE Alternative splicing of the p53 tumor suppressor gene in the Molt-4

T-lymphoblastic leukemia cell line
JOURNAL Cancer Lett. 73 (2-3), 141-148 (1993)
MEDLINE 94036762

REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbs 139316] from the original journal article.
This sequence comes from Fig. 2.

Map location: 17.
Location/Qualifiers

source 1. .160
/organism="Homo sapiens"
/db_xref="taxon:9606"

gene 1. .160
partial
/gene="p53"

CDS /note="tumor suppressor"
1. .48
partial
/gene="p53"

/note="tumor suppressor; This sequence comes from Fig. 2"
/codon_start=1
/protein_id="AAB28601.1"

/db_xref="GI:436293"
/translation="YFTIQDQTSFQKENC"

BASE COUNT 49 a 30 c 32 g 49 t
ORIGIN

Query Match 100.0%; Score 33; DB 9; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcaagtagcatctgtatcaggcaagtcataag 33
Db 100 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 68

RESULT 3

LOCUS AB018045/c

DEFINITION Homo sapiens HSP70-1 gene for heat shock protein 72, spliced
variant, partial cds.

ACCESSION AB018045
VERSION AB018045.1 GI:4691417

KEYWORDS HSP70-1; heat shock protein 72; HSP70-Hom; alternative splicing.
SOURCE Homo sapiens DNA.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)

AUTHORS Shimizu,S., Nomura,K., Ujihara,M. and Demura,H.
TITLE An additional exon of stress-inducible heat shock protein 70 gene
(HSP70-1)

JOURNAL Biochem. Biophys. Res. Commun. 257 (1), 193-198 (1999)
MEDLINE 99194576

REFERENCE 2 (bases 1 to 4360)
AUTHORS Nomura,K. and Shimizu,S.

TITLE Direct Submission
JOURNAL Submitted (27-SEP-1998) Kasru Nomura, Tokyo Women's Medical

University, Department of Medicine 2; 8-1 Kawadacho, Shinjyuku-ku,
Tokyo 162-8666, Japan {E-mail:nomura@eparkcity.ne.jp,
Tel:81-3-3353-8111(ex.39223), Fax:81-3-3357-6475)
COMMENT Sequence updated (26-Oct-1998).

FEATURES

Source 1. 4360
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="6"
/map="6p21.3"

misc_feature complement(1..196)

gene /gene="HSP70-Hom"

exon complement(1..196)
2323..2679
/gene="HSP70-Hom"

gene /gene="HSP70-1"
/note="alternative splicing"
/number=1

CDS 2323..4360
/gene="HSP70-1"
join(2632..2679,3955..>4360)

/gene="HSP70-1"
/note="spliced variant"

/codon_start=1
/product="heat shock protein 72"

exon /protein_id="BAA77235.1"
/db_xref="GI:4691418"

/translation="MKHWPQVINDGPKVQVSYKGTKAFYPEEISSMVLTKMKRI
AEAYLGYPVTNAVITVPAYFNDSQRATKDAGVIAGLNLRIINEPTAAAIAYGLDRT
GKERNVLIFDLGGCTFDVSILTIDDDGIFEVKATAGTDLGGEDFNQ"

3443..3954
/gene="HSP70-1"

/note="alternative splicing
transcription usually starts from exon 2"

exon /number=2
3955..>4360
/gene="HSP70-1"

/note="alternative splicing"
/number=3

BASE COUNT 970 a 1242 c 1147 g 1001 t
ORIGIN

Query Match 100.0%; Score 33; DB 9; Length 4360;
Best Local Similarity 100.0%; Pred. NO. 0.0014;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcaagtagcatctgtatcaggcaagtcataag 33
Db 2017 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 1985

```

RESULT      4
HSP53G/c
LOCUS
DEFINITION  HSP53G 20303 bp DNA linear PRI 25-JUN-1997
              Human p53 gene for transformation related protein p53 (also called
              transformation-associated protein p53, cellular tumor antigen p53,
              and non-viral tumour antigen p53).
X54156
ACCESSION   X54156.1 GI:35213
VERSION     1
KEYWORDS    anti-oncogene; cell cycle control; growth suppressor; heat shock
              protein 70; oncogene; p53 cellular tumour antigen; p53 gene;
              phosphoprotein; transforming capacity; tumor antigen.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 20303)
AUTHORS     Chumakov, P.M.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-1990) Chumakov P.M., Engelhardt Inst. of
              Molecular Biology, Academy of Science of the USSR, Vavilov St. 32,
              117984 Moscow, USSR
REFERENCE   2 (bases 1 to 20303)
AUTHORS     Chumakov, P.M., Almazov, V.P. and Jenkins, J.R.
TITLE       Unpublished
JOURNAL
REFERENCE   3 (bases 1 to 20303)
AUTHORS     Futreal, P.A., Barrett, J.C. and Wiseman, R.W.
TITLE       An Alu polymorphism intragenic to the TP53 gene
JOURNAL     Nucleic Acids Res. 19 (24), 6977 (1991)
MEDLINE     92107726
COMMENT     See also entries K03199, M14690, M14695, X01405, X02469, M22881-4,
              M22887-8, M22894-8.
              See also Mol. Cell. Biol. 6:1379-1385(1986);
              and Mol. Cell. Biol. 7:961-963(1987).
FEATURES    source
              1..20303
              /organism="Homo sapiens"
              /strain="caucasian"
              /db_xref="taxon:9606"
              /chromosome="17"
              /map="p13"
              join(843..949,11689..11790,11906..11927,12021..12299,
              13055..13238,13320..13432,14000..14109,14452..14588,
              14681..14754,17572..17678,18599..19876)
              /gene="p53"
prim_transcript 843..19876
              /gene="p53"
exon            843..949
              /gene="p53"
              /number=1
gene            843..19876
              /gene="p53"
intron          950..11688
              /gene="p53"
              /number=1
repeat_unit    2581..2587
              /gene="p53"
repeat_unit    2588..2877
              /note="5'-ALU flanking"
              /gene="p53"
repeat_unit    2890..2896
              /rpt_family="ALU"
              /gene="p53"
repeat_unit    3915..3929
              /note="3' ALU-flanking"
              /gene="p53"
repeat_unit    3950..4223
              /note="3' ALU-flanking"
              /gene="p53"
repeat_unit    4224..4238
              /rpt_family="ALU"
              /gene="p53"
              /translation="MEEPQSDPSVEPPLSQETFSDLWKLLENVYLSPLPSQAMDDLM"
              /db_xref="GI:35214"
              /protein_id="CAA38095.1"
              /product="protein p53"
              /codon_start=1
              /number=2
              join(11717..11790,11906..11927,12021..12299,13055..13238,
              13320..13432,14000..14109,14452..14588,14681..14754,
              17572..17678,18599..18680)
              /gene="p53"
              /codon_start=1
              /product="protein p53"
              /protein_id="CAA38095.1"
              /db_xref="GI:35214"
              /translation="MEEPQSDPSVEPPLSQETFSDLWKLLENVYLSPLPSQAMDDLM"

```



```
/number=2
join(11717..11790,11906..11927,12021..12299,13055..13238,
13320..13432,14000..14109,14452..14588,14681..14754,
17572..17678,18599..18680)
/gene="TP53"
/codon_start=1
/product="p53"
/protein_id="AAC12971.1"
/db_xref="GI:3041867"
/translation="MEEPQSDPSVEPLPSQETSDLWKLLENVLSPLPSQAMDLM
LSPDDIQWETEDCPDEARMEPAEAPVAPAPAPAPAPAPAPAPAPAPAPAPAP
YQSGYGRFLGSHGTAQVTCITYSPALNKMFCQAKTCVQLWVDSTPPGTRVRAM
AIYKQSHQMEVVRCPHRCSDSDGLAPPOHLIRVEGNLRVEYLDNRNFRHVVV
PYEPYVGSQCTTTHYMYCNSCMGMNRPILTIITLEDSSGNLIGRNSFEVRVCA
CPGRDRTEENLRKKGEPHLELPPGSTKRALPNNTSSSPQKKKPLDGYEFTLQIRG
REREFEMRELNEALELDAQAGKEPGGSRHSHLKKQGSISRHKKLMFKTEGDS
D"
variation
11827
/gene="Tp53"
/notes="Polymorphism: G or C"
/replace="c"
11906..11927
/gene="Tp53"
/number=3
11948..11951
/gene="Tp53"
/notes="GGGG in this sequence; GGG found in GenBank
Accession Number X54156"
/citation=[2]
/replace="ggg"
11952..11967
/gene="Tp53"
/notes="Polymorphism: repeat polymorphism of the
(ACCTGGAGGCTGGGG)n motif (n=1,2)"
11964..11969
/gene="Tp53"
/notes="GGGGGG in this sequence; GGGGGG found in GenBank
Accession Number X54156"
/citation=[2]
/replace="ggggggg"
12021..12299
/gene="Tp53"
/number=4
12032
/gene="Tp53"
/notes="Polymorphism: G or A"
/replace="a"
12139
/gene="Tp53"
/notes="Polymorphism: G or C"
/replace="c"
13055..13238
/gene="Tp53"
/number=5
13320..13432
/gene="Tp53"
/number=6
13399
/gene="Tp53"
/notes="Polymorphism: A or G"
/replace="g"
14000..14109
/gene="Tp53"
/number=7
14452..14588
/gene="Tp53"
/number=8
14681..14754
/gene="Tp53"
/number=9
17536
/gene="Tp53"
/notes="T in this sequence; A found in GenBank Accession
```

```
Number X54156"
/citation=[2]
/replace="a"
17572..17678
/gene="TP53"
/number=10
17699
/gene="Tp53"
/notes="C in this sequence; A found in GenBank Accession
Number X54156"
/citation=[2]
/replace="a"
18599..19876
/gene="TP53"
/number=11
18717
/gene="TP53"
/notes="A in this sequence; T found in GenBank Accession
Number X54156"
/citation=[2]
/replace="t"
BASE COUNT      5245 a      4967 c      5107 g      4984 t
ORIGIN

Query Match      100.0%; Score 33; DB 9; Length 20303;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcagtagcatctgtatcagcgaagtcag 33
|||||
Db 15034 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 15002

RESULT 6
AC087388/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-199F11 map 17, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
AC087388
AC087388.4 GI:16117576
VERSION
HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 159021)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
Homo sapiens chromosome 17, clone RP11-199F11
REFERENCE
2 (bases 1 to 159021)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lakoque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnéz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
```

JOURNAL Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 13, 2001 this sequence version replaced gi:16041379.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11969
Center clone name: 199_F_11

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 5823 5922: contig of 5822 bp in length
* 5923 5923: gap of 100 bp
* 43818 43917: contig of 37895 bp in length
* 43918 43917: gap of 100 bp
* 119707 119706: contig of 75789 bp in length
* 119707 119806: gap of 100 bp
* 119807 153208: contig of 33402 bp in length
* 153209 153308: gap of 100 bp
* 153309 159021: contig of 5713 bp in length.
* 153309 Location/Qualifiers
1. .159021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="17"
/map="17"
/clone_lib="RP11-199F11"
/clone_lib="RP11-199F11 Human Male BAC"
40373 a 39760 c 38883 g 39366 t 639 others
BASE COUNT
ORIGIN
Query Match 100.0%; Score 33; DB 2; Length 159021;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtcaagtagcatctgtatcaggcaagtcataag 33
|||||
Db 80368 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 80336
|||||
RESULT 7
AC008049/c
LOCUS
DEFINITION Homo sapiens clone RP11-199F11, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
AC008049 160457 bp DNA linear HTG 17-JUL-2001
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160457)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-olsman,F.R., Allen,C.,
Alsbrooks,S.L., Anarstange,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bivage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Somaite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 160457)
Worley,K.C.
Direct Submission
Submitted (16-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 17, 2001 this sequence version replaced gi:14328991.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMGY
Center clone name: RP11-199F11
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 66% of reads
Chemistry: Dye-terminator Big Dye; version 0.990329
Assembly program: Phrap; version 0.990329
Consensus quality: 161924 bases at least Q40
Consensus quality: 166972 bases at least Q30
Consensus quality: 169776 bases at least Q20
Estimated insert size: 162538; sum-of-contigs estimation
Quality coverage: 10.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 55987: contig of 55987 bp in length
* 55988 56087: gap of unknown length

* 56088 96796: contig of 40709 bp in length
* 96797 96896: gap of unknown length
* 96897 124682: contig of 27786 bp in length
* 124683 124782: gap of unknown length
* 124783 136186: contig of 11404 bp in length
* 136187 136286: gap of unknown length
* 136287 141789: contig of 5503 bp in length
* 141790 141889: gap of unknown length
* 141890 149503: contig of 7614 bp in length
* 149504 149604: gap of unknown length
* 149605 154667: contig of 5064 bp in length
* 154668 154767: gap of unknown length
* 154768 156994: contig of 2227 bp in length
* 156995 157094: gap of unknown length
* 157095 160457: contig of 3363 bp in length.

FEATURES
source
1..160457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-199F11"

BASE COUNT 39372 a 39284 c 40113 g 40841 t 847 others
ORIGIN

Query Match 100.0%; Score 33; DB 2; Length 160457;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtcagtagcatctgtatcaggcaagtcataag 33
|||||
Db 74697 GTCAAGTAGCATCTGTATCAGGCAAGTCATAG 74665

RESULT 8

AC087152 118985 bp DNA linear HTG 09-DEC-2000
LOCUS
DEFINITION Mus musculus clone RP23-328G13, WORKING DRAFT SEQUENCE, 24
unordered pieces.

AC087152

AC087152.1 GI:11610878

HTG: HTGS_PHASE1; HTGS_DRAFT.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 118985)

DOE Joint Genome Institute.

Sequencing of Mouse

Unpublished

2 (bases 1 to 118985)

DOE Joint Genome Institute.

Direct Submission

Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 1862981

Center clone name: RPCI-23_328G13

Summary Statistics

Consensus quality: 106849 bases at least Q40

Consensus quality: 110506 bases at least Q30

Consensus quality: 113000 bases at least Q20

Estimated insert size: 204000; agarose-fp estimation

Estimated insert size: 116685; sum-of-contigs estimation

Quality coverage: 8.1 in Q20 bases; agarose-fp estimation

Quality coverage: 14.16 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 24 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1182: contig of 1182 bp in length
* 1183 1282: gap of unknown length
* 1283 2380: contig of 1098 bp in length
* 2381 2480: gap of unknown length
* 2481 3665: contig of 1185 bp in length
* 3666 3765: gap of unknown length
* 3766 5004: contig of 1239 bp in length
* 5005 5104: gap of unknown length
* 5105 6245: contig of 1141 bp in length
* 6246 6345: gap of unknown length
* 6346 7434: contig of 1089 bp in length
* 7435 7534: gap of unknown length
* 7535 8923: contig of 1389 bp in length
* 8924 9023: gap of unknown length
* 9024 10535: contig of 1512 bp in length
* 10536 10635: gap of unknown length
* 10636 12177: contig of 1541 bp in length
* 12177 12276: gap of unknown length
* 12277 13432: contig of 1156 bp in length
* 13433 13532: gap of unknown length
* 13533 15422: contig of 1890 bp in length
* 15423 15522: gap of unknown length
* 15523 18294: contig of 2772 bp in length
* 18295 18394: gap of unknown length
* 18395 21594: contig of 3200 bp in length
* 21595 21694: gap of unknown length
* 21695 24609: contig of 2915 bp in length
* 24610 24709: gap of unknown length
* 24710 26359: contig of 1650 bp in length
* 26360 26459: gap of unknown length
* 26460 29583: contig of 3124 bp in length
* 29584 29683: gap of unknown length
* 29684 35143: contig of 5460 bp in length
* 35144 35243: gap of unknown length
* 35244 43056: contig of 7813 bp in length
* 43057 43156: gap of unknown length
* 43157 51979: contig of 8823 bp in length
* 51980 52079: gap of unknown length
* 52080 59872: contig of 7793 bp in length
* 59873 59972: gap of unknown length
* 59973 67616: contig of 7644 bp in length
* 67617 67716: gap of unknown length
* 67717 83384: contig of 15668 bp in length
* 83385 83484: gap of unknown length
* 83485 100026: contig of 16542 bp in length
* 100027 100127: gap of unknown length
* 100127 118985: contig of 18859 bp in length.

FEATURES

Location/Qualifiers

1..118985

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="RP23-328G13"

BASE COUNT 33979 a 23873 c 23761 g 35059 t 2313 others
ORIGIN

Query Match 69.7%; Score 23; DB 2; Length 118985;
Best Local Similarity 83.9%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gtcagtagcatctgtatcaggcaagtcataag 31

|||||

Db 35659 GACAGTAGTATCTAGTGGCAAAATCAT 35689

RESULT 9

```

AC027342/C
LOCUS       AC027342             112904 bp    DNA        linear    PRI 15-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2294013, complete sequence.
ACCESSION   AC027342
VERSION     AC027342.5   GI:15187263
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 112904)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 112904)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 112904)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (17-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 112904)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 15, 2001 this sequence version replaced gi:12957688.
             Draft Sequence Produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing Completed at Stanford Human Genome Center
             www.shgc.stanford.edu
             Quality: Phrap Quality >=40 99.5% of Sequence;
             Estimated Total Number of Errors is 0.3.
             STS Content:
             SHGC-3054 G13961
             WI-10043 G11676
             SHGC-36451 G30189.

FEATURES             Location/Qualifiers
     source           1..112904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2294013"
BASE COUNT          32579 a 23799 c 23407 g 33119 t
ORIGIN
Query Match        66.7%; Score 22; DB 9; Length 112904;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 aagtagcatctgtatcaggcaaaagtcatag 33
|||||
Db 19149 ATGCAGTATCTGTACACAGGCAAAATTCATAG 19120

RESULT 10
CNS01DUO/C
LOCUS       CNS01DUO             159524 bp    DNA        linear    PRI 23-MAY-2001
DEFINITION Human chromosome 14 DNA sequence BAC R-66M11 of library RPCI-11
             from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL133305
VERSION     AL133305.3   GI:14268348
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 159524)
AUTHORS     Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
             Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
             Levy,M., Eckenberg,R., Bruls,T., Deberardinis,V., Cruaud,C.,
             Gyapay,G., Saurin,W. and Weissbach,J.
             Sequencing of the human chromosome 14
             Unpublished
             2 (bases 1 to 159524)
             Genoscope.
             Direct Submission
             Submitted (23-MAY-2001) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
             On May 30, 2001 this sequence version replaced gi:7981623.
             ----- Genoscope
             Center: Genoscope / Centre National de Sequencage
             Center code: GS
             Web site: http://www.genoscope.cns.fr/
             Contact: seqref@genoscope.cns.fr

             -----
             The following BAC sequence is oriented from the T7 to the SP6 end.
             Upstream BAC (overlapping the T7 end) : C-2008K18
             Downstream BAC (overlapping the SP6 end) : R-1078I14 (AC-AL161851)
             ----- Summary Statistics
             Assembly program: Phrap; version 2.0
             Quality coverage: 6.84x in Q20 bases; sum-of-contigs
             -----
             -----
             Overall quality chart :
             Range : bases
             0 :
             1 - 9 : 13
             10 - 19 : 140
             20 - 29 : 136
             30 - 39 : 593
             40 - 49 : 2910
             50 - 59 : 5617
             60 - 69 : 9568
             70 - 79 : 20245
             80 - 89 : 44900
             90 - 99 : 75402
             -----
             Percentage of bases with a quality value >= 40 : 99 %

FEATURES             Location/Qualifiers
     source           1..159524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-66M11"
                     /clone_lib="RPCI-11"
                     /note="matching EMBL:G04384
                     21903..22028
                     RHdb:RH53791
                     RHdb:RH3691
                     dbSTS:STS26139
                     Identified using the e-PCR software (G. Schuler)"
                     75839..75988
                     /note="matching EMBL:R44409
                     RHdb:RH53609
                     dbSTS:STS26189
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT          48294 a 30052 c 31116 g 50062 t
ORIGIN
Query Match        66.7%; Score 22; DB 9; Length 159524;
Best Local Similarity 83.3%; Pred. No. 48;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggcaaaagtcacat 31
|||||
Db 116138 TCAAGTAGCTACTGTGATCAGGCAAGTCAT 116109

RESULT 11

```


AC068682
LOCUS 184536 bp DNA linear HTG 23-SEP-2000
DEFINITION Homo sapiens clone RP11-206N2, WORKING DRAFT SEQUENCE, 37 unordered
pieces.
AC068682
VERSION AC068682.3 GI:10280868
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 184536)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choapel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollard,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfave,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Vieler,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8247861.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3802
Center clone name: 206_N_2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167659 bases at least Q40
Consensus quality: 175736 bases at least Q30
Consensus quality: 178914 bases at least Q20
Insert size: 154000; agarose-fp
Insert size: 180936; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1154: contig of 1154 bp in length
* 1155 1254: gap of 100 bp
* 1255 2766: contig of 1512 bp in length
* 2767 2866: gap of 100 bp
* 2867 4073: contig of 1207 bp in length
* 4074 4173: gap of 100 bp
* 4174 5191: contig of 1018 bp in length
* 5192 5291: gap of 100 bp
* 5292 7387: contig of 2096 bp in length
* 7388 7487: gap of 100 bp
* 7488 9087: contig of 1600 bp in length
* 9088 9187: gap of 100 bp
* 9188 10361: contig of 1174 bp in length
* 10362 10461: gap of 100 bp
* 10462 12065: contig of 1604 bp in length
* 12066 12165: gap of 100 bp
* 12166 13498: contig of 1333 bp in length
* 13499 13598: gap of 100 bp
* 13599 15085: contig of 1487 bp in length
* 15086 15185: gap of 100 bp
* 15186 17249: contig of 2064 bp in length
* 17250 17349: gap of 100 bp
* 17350 19243: contig of 1894 bp in length
* 19244 19343: gap of 100 bp
* 19344 21127: contig of 1784 bp in length
* 21128 21227: gap of 100 bp
* 21228 23345: contig of 2118 bp in length
* 23346 23445: gap of 100 bp
* 23446 25965: contig of 2520 bp in length
* 25966 26065: gap of 100 bp
* 26066 28195: contig of 2130 bp in length
* 28196 28295: gap of 100 bp
* 28296 32216: contig of 3921 bp in length
* 32217 33316: gap of 100 bp
* 33317 36954: contig of 4638 bp in length
* 36955 37054: gap of 100 bp
* 37055 40348: contig of 3294 bp in length
* 40349 40448: gap of 100 bp
* 40449 43556: contig of 3108 bp in length
* 43557 43656: gap of 100 bp
* 43657 47926: contig of 4270 bp in length
* 47927 48026: gap of 100 bp
* 48027 54079: contig of 6053 bp in length
* 54080 54179: gap of 100 bp
* 54180 58962: contig of 4783 bp in length
* 58963 59062: gap of 100 bp
* 59063 63739: contig of 4677 bp in length
* 63740 63839: gap of 100 bp
* 63840 71551: contig of 7712 bp in length
* 71552 71651: gap of 100 bp
* 71652 78218: contig of 6567 bp in length
* 78219 78318: gap of 100 bp
* 78319 84982: contig of 6664 bp in length
* 84983 85082: gap of 100 bp
* 85083 91930: contig of 6848 bp in length
* 91931 92030: gap of 100 bp
* 92031 98922: contig of 6892 bp in length
* 98923 99022: gap of 100 bp
* 99023 106496: contig of 7474 bp in length
* 106497 106596: gap of 100 bp
* 106597 114950: contig of 8354 bp in length
* 114951 115050: gap of 100 bp
* 115051 124353: contig of 9303 bp in length
* 124354 124453: gap of 100 bp
* 124454 132639: contig of 8186 bp in length
* 132640 132739: gap of 100 bp
* 132740 141572: contig of 8833 bp in length
* 141573 141672: gap of 100 bp
* 141673 151302: contig of 9630 bp in length
* 151303 151402: gap of 100 bp
* 151403 161767: contig of 10365 bp in length
* 161768 161867: gap of 100 bp
* 161868 184536: contig of 22669 bp in length.
* Location/Qualifiers

FEATURES

* NOTE: this is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.

*	1	14199:	contig of 14199 bp in length
*	14200:	gap of unknown length	
*	14299:	contig of 12096 bp in length	
*	26395:	contig of unknown length	
*	26396:	contig of 4821 bp in length	
*	26496:	gap of unknown length	
*	31317:	contig of 6900 bp in length	
*	31417:	contig of 6900 bp in length	
*	38317:	gap of unknown length	
*	38417:	contig of 5117 bp in length	
*	43534:	gap of unknown length	
*	43634:	contig of 4591 bp in length	
*	48225:	gap of unknown length	
*	48325:	contig of 4619 bp in length	
*	52943:	gap of unknown length	
*	53043:	contig of 5292 bp in length	
*	58335:	gap of unknown length	
*	58435:	contig of 5158 bp in length	
*	63593:	gap of unknown length	
*	63693:	contig of 4019 bp in length	
*	63694:	gap of unknown length	
*	67712:	contig of 4019 bp in length	
*	67713:	gap of unknown length	
*	67813:	contig of 3364 bp in length	
*	71377:	gap of unknown length	
*	71477:	contig of 4336 bp in length	
*	75813:	gap of unknown length	
*	75913:	contig of 3396 bp in length	
*	79308:	gap of unknown length	
*	79409:	contig of 3391 bp in length	
*	79409:	gap of unknown length	
*	82800:	contig of 1627 bp in length	
*	82900:	gap of unknown length	
*	84527:	contig of 2814 bp in length	
*	84627:	gap of unknown length	
*	87441:	contig of 2814 bp in length	
*	87540:	gap of unknown length	
*	90385:	contig of 2845 bp in length	
*	90386:	gap of unknown length	
*	90485:	contig of 4148 bp in length	
*	90486:	gap of unknown length	
*	94733:	contig of 2460 bp in length	
*	94734:	gap of unknown length	
*	97193:	contig of 2460 bp in length	
*	97194:	gap of unknown length	
*	97293:	contig of 2341 bp in length	
*	97294:	gap of unknown length	
*	100234:	contig of 2695 bp in length	
*	100235:	gap of unknown length	
*	100335:	contig of 2695 bp in length	
*	103023:	gap of unknown length	
*	103030:	contig of 2470 bp in length	
*	103130:	gap of unknown length	
*	103130:	contig of 2470 bp in length	
*	105600:	gap of unknown length	
*	105700:	contig of 2343 bp in length	
*	108043:	gap of unknown length	
*	108043:	contig of 2643 bp in length	
*	110783:	gap of unknown length	
*	110786:	contig of 3100 bp in length	
*	110886:	gap of unknown length	
*	113986:	contig of 3068 bp in length	
*	113986:	gap of unknown length	
*	114086:	contig of 2769 bp in length	
*	117154:	gap of unknown length	
*	117253:	contig of 2898 bp in length	
*	120022:	gap of unknown length	
*	120123:	contig of 2898 bp in length	
*	120232:	gap of unknown length	
*	123021:	contig of 1577 bp in length	
*	123212:	gap of unknown length	
*	124697:	contig of 1836 bp in length	
*	124698:	gap of unknown length	
*	126633:	contig of 2103 bp in length	
*	126734:	gap of unknown length	
*	128837:	contig of 2534 bp in length	
*	128937:	gap of unknown length	
*	131470:	contig of 1681 bp in length	
*	131570:	gap of unknown length	
*	131571:	contig of 1681 bp in length	


```
* 69944 70043: gap of unknown length
* 70044 71154: contig of 1111 bp in length
* 71154 71254: gap of unknown length
* 71254 72710: contig of 1456 bp in length
* 72710 72811: gap of unknown length
* 72811 74269: contig of 1459 bp in length
* 74269 74370: gap of unknown length
* 74370 76021: contig of 1651 bp in length
* 76021 76120: gap of unknown length
* 76120 77287: contig of 1167 bp in length
* 77287 77388: gap of unknown length
* 77388 79081: contig of 1694 bp in length
* 79081 79181: gap of unknown length
* 79181 80235: contig of 1044 bp in length
* 80235 80326: gap of unknown length
* 80326 81431: contig of 1106 bp in length
* 81431 81531: gap of unknown length
* 81531 82579: contig of 1048 bp in length
* 82579 82679: gap of unknown length
* 82679 84130: contig of 1451 bp in length
* 84130 84230: gap of unknown length
* 84230 85514: contig of 1284 bp in length
* 85514 85644: gap of unknown length
* 85644 86839: contig of 1225 bp in length
* 86839 86939: gap of unknown length
* 86939 87965: contig of 1026 bp in length
* 87965 88065: gap of unknown length
* 88065 89166: contig of 1101 bp in length
* 89166 89266: gap of unknown length
* 89266 90620: contig of 1354 bp in length
* 90620 90720: gap of unknown length
* 90720 91767: contig of 1047 bp in length
* 91767 91867: gap of unknown length
* 91867 93225: contig of 1358 bp in length
* 93225 93326: gap of unknown length
* 93326 94683: contig of 1358 bp in length
* 94683 94783: gap of unknown length
* 94783 96476: contig of 1693 bp in length
* 96476 96576: gap of unknown length
* 96576 97643: contig of 1067 bp in length
* 97643 97743: gap of unknown length
* 97743 97644
```

Query Match Best Local Similarity 66.1%; Score 21.8; DB 2; Length 98779;

Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 gtcagtagcatctgtatccggcaagtcataag 33

Db 64088 GTCAAGTGCATCTGAATCCCCCAGAGTCANAG 64120

RESULT 14

AL390245
LOCUS Homo sapiens chromosome 1 clone RP4-794P10, linear HTG 13-NOV-2001
DEFINITION PROGRESS ***, 11 unordered pieces.

ACCESSION AL390245

VERSION AL390245.3 GI:9909482

KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

Mclay, K.

Direct Submission

TITLE Submitted (06-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,

JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 25, 2000 this sequence version replaced gi:9798332.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj794P10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 48% of reads Chemistry:
Dye-terminator Big Dye; 51% of reads
Consensus quality: 139550 bases at least Q40
Consensus quality: 141905 bases at least Q30
Consensus quality: 143156 bases at least Q20
Insert size: 14411; sum-of-contigs
Insert size: 136639; 7.1% error; agarose-fp
Quality coverage: 4.14x in Q20 bases; sum-of-contigs Quality
coverage: 4.36x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 12532: contig of 12532 bp in length
* 12533 12632: gap of 100 bp
* 12633 22864: contig of 10232 bp in length
* 22865 22964: gap of 100 bp
* 22965 25731: contig of 2767 bp in length
* 25732 25831: gap of 100 bp
* 25832 29385: contig of 3554 bp in length
* 29386 29485: gap of 100 bp
* 29486 32618: contig of 3133 bp in length
* 32619 32718: gap of 100 bp
* 32719 34850: contig of 2132 bp in length
* 34851 34950: gap of 100 bp
* 34951 41868: contig of 6918 bp in length
* 41869 41968: gap of 100 bp
* 41969 46297: contig of 4329 bp in length
* 46298 46397: gap of 100 bp
* 46398 97392: contig of 50995 bp in length
* 97393 97492: gap of 100 bp
* 97493 105227: contig of 7735 bp in length
* 105228 105327: gap of 100 bp
* 105328 145111: contig of 39784 bp in length.

FEATURES

source

1..145111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-794P10"
/clone_lib="RPCI-4"

misc_feature

1..12532
/note="assembly_fragment:00313
fragment_chain:1"
clone_end:SP6
vector_side:left"
12633..22864

misc_feature

/note="assembly_fragment:00680
fragment_chain:1"
22965..25731
/note="assembly_fragment:00337
fragment_chain:2"

misc_feature

25832..29385
/note="assembly_fragment:00524
fragment_chain:2"

misc_feature

29486..32618
/note="assembly_fragment:00102
fragment_chain:2"

misc_feature

32719..34850
/note="assembly_fragment:00306"

misc_feature

34951..41868

```
/note="assembly_fragment:00503"
41969. .46297
/note="assembly_fragment:00559"
46398. .97392
/note="assembly_fragment:00658"
97493. 105227
/note="assembly_fragment:00867"
105328. .145111
/note="assembly_fragment:01072"
BASE COUNT 42567 a 28443 c 29392 g 43704 t 1005 others
ORIGIN

Query Match 66.1%; Score 21.8; DB 2; Length 145111;
Best Local Similarity 78.8%; Pred. NO. 59;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaagtcataag 33
|||||
Db 89296 GTCAAGTAGCATCTATAAATGTCAAATTCAAAG 89328

RESULT 15
AL645507/C 202174 bp DNA linear HTG 05-DEC-2001
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-476C24, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.
ACCESSION AL645507
VERSION AL645507.6 GI:16944311
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Direct Submission
Submitted (03-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 15, 2001 this sequence version replaced gi:16904503.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba476C24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 200245 bases at least Q40
Consensus quality: 200762 bases at least Q30
Consensus quality: 201144 bases at least Q20
Insert size: 201574; sum-of-contigs
Insert size: 205408; 11.6% error; agarose-fp
Quality coverage: 6.66x in Q20 bases; sum-of-contigs Quality
coverage: 6.53x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 100046: contig of 100046 bp in length
* 100047 100146: gap of 100 bp
* 100147 100313: contig of 3167 bp in length
* 103314 103413: gap of 100 bp
* 103414 156103: contig of 52690 bp in length
```

```
* 156104 156203: gap of 100 bp
* 156204 184218: contig of 28015 bp in length
* 184219 184318: gap of 100 bp
* 184319 187120: contig of 2802 bp in length
* 187121 187220: gap of 100 bp
* 187221 199827: contig of 12607 bp in length
* 199828 199927: gap of 100 bp
* 199928 202174: contig of 2247 bp in length.
FEATURES
Location/Qualifiers
source
1. 202174
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-476C24"
/clone_lib="RPC1-11.2"
1. 100046
/note="assembly_fragment:00136"
fragment_chain:1
clone_end:SP6
vector_side:left"
100147. 103313
/note="assembly_fragment:01717"
fragment_chain:1"
103414. 156103
/note="assembly_fragment:02913"
fragment_chain:1"
156204. 184218
/note="assembly_fragment:02625"
fragment_chain:1"
184319. 187120
/note="assembly_fragment:02061"
fragment_chain:1"
187221. 199827
/note="assembly_fragment:01216"
fragment_chain:1"
199928. 202174
/note="assembly_fragment:00448"
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 62225 a 39698 c 40511 g 59139 t 601 others
ORIGIN

Query Match 66.1%; Score 21.8; DB 2; Length 202174;
Best Local Similarity 78.8%; Pred. No. 58;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaagtcataag 33
|||||
Db 111808 GTCAAGTAGCATCTATAAATGTCAAATTCAAAG 111776

Search completed: August 24, 2002, 21:54:01
Job time: 15526 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 22:11:51 ; Search time 905.65 Seconds
(without alignments)
62.561 Million cell updates/sec

Title: US-09-986-381-3
Perfect score: 33
Sequence: 1 gtaacagtagcatctgtatcaggcaagtcataag 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-n-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	20.4	61.8	359	22 AAL09224	Human breast cancer
C 2	20.4	61.8	371	22 AAL17113	Human breast cancer
C 3	20.4	61.8	726	22 AAL197198	Human neuroblastoma
4	20.4	61.8	1610	20 AAL42141	Human normal bladder
5	20.4	61.8	1619	21 AAF16311	Human prostate cancer
6	20.4	61.8	1975	22 AAF44936	cDNA encoding novel
7	19.8	60.0	735	22 AAI96659	Human neuroblastoma
8	19.8	60.0	4721	23 AAS84613	DNA encoding novel
9	19.4	58.8	449	22 AAL28850	Drosophila melanog

10	19.4	58.8	1045	18 AAT59977	5' untranslated re
11	19.4	58.8	1045	20 AAX02016	D. melanogaster ti
C 12	19.4	58.8	1529	21 AAC54931	Arabidopsis thalia
C 13	19.4	58.8	1539	21 AAC53145	Arabidopsis thalia
14	19.4	58.8	3664	23 ABL02242	Drosophila melanog
15	19.4	58.8	3954	18 AAT59975	Full length tipe p
16	19.4	58.8	3954	20 AAX02015	D. melanogaster ti
C 17	19.2	58.2	73	17 AAT12383	DNA encoding human
C 18	19.2	58.2	73	20 AAV84086	HRV-14 epitope ins
C 19	19.2	58.2	300	14 AAX05905	Human brain Exptes
C 20	19.2	58.2	387	20 AAX56535	Human AAL52150 DNA
C 21	19.2	58.2	441	22 AAS34232	Human cDNA encodin
C 22	19.2	58.2	467	20 AAX56542	Human N28398 DNA f
C 23	19.2	58.2	506	20 AAV88347	EST clone GB814.
24	19.2	58.2	1305	20 AAX61747	B. burgdorferi ant
25	19.2	58.2	1486	20 AAX02855	Human zsig46 DNA.
C 26	19.2	58.2	1663	21 AAC43886	Arabidopsis thalia
C 27	19.2	58.2	1751	20 AAZ24826	Human secreted pro
C 28	19.2	58.2	1857	23 AAS53130	Enterococcus faeca
C 29	19.2	58.2	2120	22 AAK94829	Human full-length
C 30	19.2	58.2	2241	22 AAK94509	Human P21-active k
31	19.2	58.2	3932	22 AAH44557	Human OREF ORF1509
C 32	19.2	58.2	4796	21 AAX75954	Enterococcus faeca
C 33	19.2	58.2	5830	20 AAX13218	Human polynucleoti
C 34	19.2	58.2	910715	20 AAX20248	Borrelia burgdorfe
35	19	57.6	4378	22 AAI60050	Human polynucleoti
36	19	57.6	5102	22 AAL58264	Human musculoskele
37	19	57.6	13444	22 AAL37295	Human reproducing
C 38	18.8	57.0	451	22 AAL01243	cDNA encoding nove
39	18.8	57.0	1155	22 AAS44935	cDNA encoding nove
40	18.8	57.0	2095	22 AAS45123	cDNA encoding nove
C 41	18.8	57.0	2095	22 AAS45124	DNA encoding novel
C 42	18.8	57.0	13960	23 AAS87164	Human cDNA 5'-end
43	18.6	56.4	697	22 AAK92158	Human cDNA clone r
C 44	18.6	56.4	697	22 AAK93424	KIP1 coding sequen
C 45	18.6	56.4	5093	16 AAX99876	

ALIGNMENTS

RESULT 1
AAL09224/C
ID AAL09224 standard; cDNA; 359 BP.
XX
AC AAL09224;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 1681.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX

```
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 337; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 359 BP; 107 A; 70 C; 70 G; 112 T; 0 other;

Query Match 61.8%; Score 20.4; DB 22; Length 359;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 caagtagcatctgtatcaggcaaatcata 32
   ||| | ||||| ||||| ||| | |
Db 252 CAATTGTCATCTGTATCAGGCATAGAGAGA 223

RESULT 2
AAL17113/C
ID AAL17113 standard; cDNA; 371 BP.
XX
AC AAL17113;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 9570.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PS 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 1707; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 371 BP; 119 A; 69 C; 62 G; 121 T; 0 other;

Query Match 61.8%; Score 20.4; DB 22; Length 371;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 caagtagcatctgtatcaggcaaatcata 32
   ||| | ||||| ||||| ||| | |
Db 237 CAATTGTCATCTGTATCAGGCATAGAGAGA 208

RESULT 3
AAI97198/G
ID AAI97198 standard; cDNA; 726 BP.
XX
AC AAI97198;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3273.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
XX
PI (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
DR WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents
XX
PS Claim 1; Page 2377-2378; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 726 BP; 191 A; 144 C; 115 G; 256 T; 20 other;

Query Match 61.8%; Score 20.4; DB 22; Length 726;
Best Local Similarity 95.5%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tcaagtagcatctgtatcaggc 23
   ||||| ||||| |||||
Db 484 TCAAGTAACATCTGTATCAGGC 463

RESULT 4
AAZ42141
ID AAZ42141 standard; cDNA; 1610 BP.
```

XX AAZ42141;
AC 31-JAN-2000 (first entry)
XX Human normal bladder tissue cDNA derived EST 20.
DE Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
XX cancer; gene therapy; ss.
KW Homo sapiens.
OS DE19818620-A1.
PN 28-OCT-1999.
PD 21-APR-1998; 98DE-1018620.
PF 21-APR-1998; 98DE-1018620.
PR (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
PA WPI: 1999-602416/52.
PI New polypeptides and their nucleic acids, useful for treatment of
XX bladder tumour and identification of therapeutic agents -
PS Claim 3; Page 163; 366pp; German.
XX This invention describes novel polypeptide fragment sequences (I) and
CC their encoding nucleic acids (II) which are highly expressed in normal
CC bladder tissue and have cytostatic activity. (II) are used for
CC recombinant expression of (I) and to isolate complete genes. (I) are
CC used to identify agents suitable for the treatment of bladder tumours,
CC to directly treat this form of cancer (including expression from gene
CC therapy vectors), or are used in a preparation for cancer treatment. (I)
CC is also used for the generation of specific antibodies. (II) are
CC identified by assembling ESTs (expressed sequence tags) from a
CC particular tissue type before comparison of expression patterns. This
CC allows a significantly longer fragment of the gene to be revealed, and
CC therefore reduces the number of failures because of ESTs from different
CC libraries representing different parts of the same unknown gene
CC distorting the estimated frequency of occurrence in a particular tissue.
CC AAZ42122-742248 represent EST fragments derived from a human normal
CC bladder tissue cDNA library which encode the protein fragments
CC represented in AAY60329-Y60591.
XX Sequence 1610 BP; 469 A; 320 C; 412 G; 409 T; 0 other;
SQ

Query Match 61.8%; Score 20.4; DB 20; Length 1610;
Best Local Similarity 95.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggc 23
||||| |||||||||
Db 1150 tcaagtaacatctgtatcaggc 1171

RESULT 5
AAF16311
ID AAF16311 standard; cDNA; 1619 BP.
XX AAF16311;
AC 13-MAR-2001 (first entry)
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:746.
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

KW vulnery; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease; ss.
XX Homo sapiens.
OS WO2000055174-A1.
PN 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05988.
PF 12-MAR-1999; 99US-0124270.
PR (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM;
XX WPI: 2000-587513/55.
DR P-PSDB; AAB57108.
DR Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX Claim 1; Page 1168-1169; 2338pp; English.
PS AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX Sequence 1619 BP; 495 A; 314 C; 402 G; 405 T; 3 other;
SQ

Query Match 61.8%; Score 20.4; DB 21; Length 1619;
Best Local Similarity 95.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggc 23
||||| |||||||||
Db 1125 tcaagtaacatctgtatcaggc 1146

RESULT 6
AAS44936
ID AAS44936 standard; cDNA; 1975 BP.
XX AAS44936;
AC 18-DEC-2001 (first entry)
XX CDNA encoding novel human secretory protein, Seq ID No 17.
DE Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
XX ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;

KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
 KW fertility; analgesic; pain; antigen; ss.
 XX Homo sapiens.
 OS
 PN WO200166689-A2.
 XX
 PD 13-SEP-2001.
 XX
 XX 05-MAR-2001; 2001WO-US04942.
 PF
 XX 07-MAR-2000; 2000US-0519705.
 PR 19-MAY-2000; 2000US-0574454.
 PR 17-JUN-2000; 2000US-0596193.
 PR 14-JUL-2000; 2000US-0616847.
 PR 19-SEP-2000; 2000US-0665363.
 PR 20-OCT-2000; 2000US-0693267.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
 PI Zhao QA, Yang Y, Dmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
 XX
 DR P-PSDB; AAU28036.
 DR
 XX Novel polypeptides and nucleic acids obtained from cDNA libraries
 PT prepared from various human tissues, for diagnosis and treatment of
 PT cancer, neurological, inflammatory, and autoimmune disorders -
 XX
 XX Claim 1; SEQ ID No 17; 107pp: English.
 PS
 XX The invention relates to novel isolated human secreted polypeptides (I)
 CC and polynucleotides (II). (I) and (II) are useful for treating
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
 CC involved in increasing haematopoiesis, stem cell survival, bone growth
 CC and remodeling. (I), (II) and modulators of (II) are useful for
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for
 CC creating transgenic animals useful for studying the in vivo activities of
 CC the polypeptide as well as for studying modulators of the polypeptides.
 CC (I) induces the proliferation of neural cells and regeneration of nerve
 CC and brain tissue and is useful for the treatment of central and
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
 CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth, and in tissue repair, healing of burns, incisions,
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
 CC disorders, or periodontal disease. Furthermore, (I) is also useful for
 CC gut protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, various immune deficiencies and
 CC disorders including severe combined immunodeficiency (SCID), bacterial or
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 CC reactions and conditions, such as asthma or other respiratory problems.
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 CC analgesic effects or other pain reducing effects, immunoglobulin like
 CC activity and can act as an antigen in a vaccine composition to raise an
 CC immune response. AAS44920-AAS45295 represent novel human secreted protein
 CC coding sequences of the invention.
 XX
 XX Sequence 1975 BP; 575 A; 366 C; 466 G; 568 T; 0 other;
 SQ
 Query Match 61.8%; Score 20.4; DB 22; Length 1975;
 Best Local Similarity 95.5%; Pred. No. 26;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcagc 23
 Db 1102 tcaagtagcatctgtatcagc 1123
 RESULT 7
 AAI96659
 ID AAI96659 standard; cDNA; 735 BP.
 XX
 AC AAI96659;
 XX
 DT 13-NOV-2001 (first entry)
 XX
 DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2734.
 XX
 XX Human; neuroblastoma; malignancy; Cancer; tumour marker; N-myc; TrkA; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200166719-A1.
 XX
 PD 13-SEP-2001.
 XX
 PF 02-MAR-2001; 2001WO-JP01629.
 XX
 PR 07-MAR-2000; 2000JP-0159195.
 XX
 PA (CHIB-) CHIBA PREFECTURE.
 PA (HISM) HISAMITSU PHARM CO LTD.
 XX
 XX Nakagawara A;
 PI
 XX WPI; 2001-565584/63.
 DR
 XX Nucleic acids originating in gene expressed in human neuroblastoma,
 PT useful as probe or primer in diagnosing prognosis of human
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
 PT for anti-cancer agents -
 XX
 XX Claim 1; Page 2004; 2979pp: Japanese.
 PS
 XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
 CC human neuroblastoma. The nucleic acids are applicable as a probe or
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
 CC susceptibility indicators or tumour markers for anti-cancer agents. The
 CC gene information for diagnosing prognosis is related to factors similar
 CC to that for N-myc and TrkA genes.
 XX
 XX Sequence 735 BP; 223 A; 110 C; 135 G; 251 T; 16 other;
 SQ
 Query Match 60.0%; Score 19.8; DB 22; Length 735;
 Best Local Similarity 77.4%; Pred. No. 40;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 Qy 3 caagtagcatctgtatcaggcaaatcag 33
 Db 71 catgtagaatctaaataggcaaatcag 101
 RESULT 8
 AAS84613
 ID AAS84613 standard; cDNA; 4721 BP.
 XX
 AC AAS84613;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #20417.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX

```
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABC20426.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID No 20417; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 4721 BP; 1417 A; 995 C; 1123 G; 1186 T; 0 other;

Query Match 60.0%; Score 19.8; DB 23; Length 4721;
Best Local Similarity 77.4%; Pred. No. 58;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 gtcagtagcatctgtatcaggcaaatcat 31
   | | | | | | | | | | | | | | | | | | | |
Db 4249 gaccagtagcaccgtgtatcattgaaactcat 4279

RESULT 9
AAH28850
ID AAH28850 standard; DNA; 449 BP.
XX AAH28850;
XX 17-JUL-2001 (first entry)
XX Drosophila melanogaster essential gene fragment, SEQ ID NO: 39.
XX Drosophila melanogaster; fruit fly; essential gene; screening assay;
XX pesticide; crop protection; chromosome 3; ds.
XX Drosophila melanogaster.

Query Match 58.8%; Score 19.4; DB 22; Length 449;
Best Local Similarity 79.3%; Pred. No. 54;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggcaaatca 30
   | | | | | | | | | | | | | | | | | | | |
Db 248 tcaagaagcagctaaatcaggcaaaagca 276

RESULT 10
AAT59977
ID AAT59977 standard; DNA; 1045 BP.
XX AAT59977;
XX 12-MAY-1997 (first entry)
XX 5' untranslated region of the tipE protein coding sequence.
XX Drosophila; tipE; para protein; voltage-dependent cation channel; stroke;
XX pesticide; insecticide; insect; parasitic infection; human; head trauma;
XX neuroprotection; hypoxia; therapy; ss.
XX Drosophila melanogaster.
XX US5593862-A.
XX 14-JAN-1997.
XX 04-OCT-1994; 94US-0317880.
XX 04-OCT-1994; 94US-0317880.
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX Feng G, Hall LM;
XX WPI; 1997-099467/09.
XX
```

```
XX WO200118547-A1.
XX 15-MAR-2001.
XX 06-SEP-2000; 2000WO-GB03444.
XX 07-SEP-1999; 99GB-0021009.
XX (UNIU ) UNIV GLASGOW.
XX Davies RW, Kaiser K, Yang MY;
XX WPI; 2001-281436/29.
XX Screening assays for used for identifying compounds having a
XX physiological effect on proteins identified as being essential -
XX Claim 1; Page 139; 695pp; English.
XX The present sequence is part of an essential gene from Drosophila
XX melanogaster. Lack of expression of the protein encoded by this
XX gene leads to a lethal or semi-lethal phenotype. The invention
XX relates to 902 nucleic acid sequences from genes encoding proteins
XX which are thought to be essential, and to a screening assay for
XX identifying compounds which have a physiological effect on these
XX proteins. Suitable compounds are useful as pesticides and may be used
XX in conjunction with other pesticides and herbicides for crop
XX protection. The gene corresponding to the present sequence is located
XX on chromosome 3.
XX Sequence 449 BP; 168 A; 105 C; 101 G; 75 T; 0 other;

Query Match 58.8%; Score 19.4; DB 22; Length 449;
Best Local Similarity 79.3%; Pred. No. 54;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggcaaatca 30
   | | | | | | | | | | | | | | | | | | | |
Db 248 tcaagaagcagctaaatcaggcaaaagca 276

RESULT 10
AAT59977
ID AAT59977 standard; DNA; 1045 BP.
XX AAT59977;
XX 12-MAY-1997 (first entry)
XX 5' untranslated region of the tipE protein coding sequence.
XX Drosophila; tipE; para protein; voltage-dependent cation channel; stroke;
XX pesticide; insecticide; insect; parasitic infection; human; head trauma;
XX neuroprotection; hypoxia; therapy; ss.
XX Drosophila melanogaster.
XX US5593862-A.
XX 14-JAN-1997.
XX 04-OCT-1994; 94US-0317880.
XX 04-OCT-1994; 94US-0317880.
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX Feng G, Hall LM;
XX WPI; 1997-099467/09.
XX
```

PT Nucleic acid encoding Drosophila melanogaster tipE protein - for
 PT prodn. of recombinant voltage-dependent cation channel
 XX
 PS Example 11; Column 35-36; 33pp; English.
 XX
 CC AAT59977 and AAT59978 represent the 5' and 3' untranslated regions of
 CC the Drosophila tipE protein (see AAM13843) coding sequence (see
 CC AAT59975). Mutations in the tipE protein, result in a ethyl methane
 CC sulphonate- induced recessive mutation phenotype. Homologous flies for
 CC the mutation paralyse rapidly at 38 degrees, and recover immediately when
 CC returned to 23 degrees. Coexpression of the full length tipE sequence,
 CC and a nucleic acid encoding a para protein results in translation
 CC products that form a functional voltage-dependent cation channel. The
 CC cation channel can be used to screen for pesticides active against
 CC insects such as Drosophila melanogaster and pest insects. The cation
 CC channel can also be used to screen for drugs for use in the treatment and
 CC prevention of parasitic infections in humans and animals, and to screen
 CC drugs for their neuroprotective effect against hypoxia, stroke, and head
 CC trauma.
 XX
 SQ Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;

Query Match 58.8%; Score 19.4; DB 18; Length 1045;
 Best Local Similarity 79.3%; Pred. No. 64;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 2 tcaagtagcatctgtatcaggcaaaagtca 30
 Db 459 tcaagaagcagctaaatcaggcaaaagca 487

RESULT 11
 AAX02016
 ID AAX02016 standard; DNA; 1045 BP.
 XX
 AC AAX02016;
 XX
 DT 21-APR-1999 (first entry)
 DE
 XX D. melanogaster tipE+ 4Kb clone 5'-UTR.
 XX
 KW tipE+; para protein; modulator; voltage dependent cation channel; VDCC;
 KW pesticide; insect control; pharmaceutical agent; neuroprotection;
 KW hypoxia; ischaemia; stroke; head trauma; ss.
 XX
 OS Drosophila melanogaster.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..1045
 FT /*tag= a
 XX
 XX US5871940-A.
 XX
 PD 16-FEB-1999.
 XX
 XX 13-JAN-1997; 97US-0782396.
 XX
 XX 04-OCT-1994; 94US-0317880.
 PR 13-JAN-1997; 97US-0782396.
 XX
 XX (UYNV) UNIV NEW YORK STATE RES FOUND.
 PA
 XX Feng G, Hall LM;
 PI
 XX WPI; 1999-166632/14.
 DR
 XX Screening for agents which modulate ion channel function - using
 PT host cells transformed with nucleic acids encoding the Drosophila
 PT melanogaster tipE and para proteins
 XX
 XX Disclosure; Column 35-36; 54pp; English.
 PS
 XX

CC This sequence represents the 5'-UTR from a Drosophila melanogaster tipE+
 CC protein which is used in a method for screening for agents which modulate
 CC ion channel function which uses host cells transformed with nucleic acid
 CC encoding Drosophila melanogaster tipE and para proteins. Co-expression of
 CC these genes in the host cell, allows the formation of a functional
 CC voltage dependent cation channel (VDCC) in the cell. The agents
 CC identified can be used as pesticides for the control of Drosophila
 CC melanogaster or other insects. They can also be used to screen
 CC pharmaceutical agents for their neuroprotective affect against e.g.
 CC hypoxia, ischaemia, stroke and head trauma.
 XX
 SQ Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;

Query Match 58.8%; Score 19.4; DB 20; Length 1045;
 Best Local Similarity 79.3%; Pred. No. 64;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 2 tcaagtagcatctgtatcaggcaaaagtca 30
 Db 459 tcaagaagcagctaaatcaggcaaaagca 487

RESULT 12
 AAC54931/C
 ID AAC54931 standard; DNA; 1529 BP.
 XX
 AC AAC54931;
 XX
 DT 18-OCT-2000 (first entry)
 DE
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 79582.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999;	99US-0134941.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137528.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 11-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139730.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.

PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
Query Match 58.8%; Score 19.4; DB 21; Length 1529;					
Best Local Similarity 79.3%; Pred. No. 69;					
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;					
QY	3	caagtagcatctgtatcaggaagtcacat 31	PR	10-JUN-1999;	99US-0138847.
			PR	14-JUN-1999;	99US-0139119.
Db	1080	CAAGTAGCTGCTCGTTGCGAAGTCCT 1052	PR	16-JUN-1999;	99US-0139452.
RESULT 13					
AAC53145/C					
ID	AAC53145 standard; DNA; 1539 BP.				
XX	AAC53145;				
AC	AC				
XX	XX				
DT	18-OCT-2000 (first entry)				
XX	XX				
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 73527.				
XX	XX				
KW	Hybridisation assay; genetic mapping; gene expression control;				
KW	protein identification; signal transduction pathway;				
KW	metabolic pathway; promoter; termination sequence; ss.				
XX	XX				
OS	Arabidopsis thaliana.				
XX	XX				
PN	EP1033405-A2.				
XX	XX				
PD	06-SEP-2000.				
XX	XX				
PF	25-FEB-2000; 2000EP-0301439.				
XX	XX				
PR	25-FEB-1999;	99US-0121825.	PR	12-JUL-1999;	99US-0142977.
PR	05-MAR-1999;	99US-0123180.	PR	13-JUL-1999;	99US-0143542.
PR	09-MAR-1999;	99US-0123548.	PR	14-JUL-1999;	99US-0143624.
PR	23-MAR-1999;	99US-0125788.	PR	15-JUL-1999;	99US-0144005.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144085.
PR	29-MAR-1999;	99US-0126785.	PR	16-JUL-1999;	99US-0144086.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144325.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144331.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144332.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144333.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144334.
PR	21-APR-1999;	99US-0130449.	PR	19-JUL-1999;	99US-0144335.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0131449.	PR	20-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	99US-0132484.	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR			PR	04-AUG-1999;	99US-0147302.


```
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-016192.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 58.8%; Score 19.4; DB 21; Length 1539;
Best Local Similarity 79.3%; Pred. No. 69;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 caagttagcgtctgcgtttggcaagtcct 31
   ||||| |||| | ||||| |||
Db 1082 CAAGTAGCGTCTGCGTTTGGCAAGTCCT 1054

RESULT 14
ABL02242
ID ABL02242 standard; cDNA; 3664 BP.
XX
AC ABL02242;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1208.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEXE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
DR P-PSDB; ABB58139.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
   genes from Drosophila and for elucidating cell signalling and cell-cell
   interactions -
XX
PS Claim 1; SEQ ID NO 1208; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
   capable of detecting 1000 or more genes from Drosophila. The invention is
   useful in developmental biology and in elucidating cell signalling and
   cell-cell interactions in higher eukaryotes for the development of
   insecticides, therapeutics and pharmaceutical drugs. The invention
   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
   sequences (ABL01840-ABL16175) and the encoded proteins
   (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
   specification, but was obtained in electronic format directly from WIPO
   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3664 BP; 1150 A; 895 C; 837 G; 782 T; 0 other;

Query Match 58.8%; Score 19.4; DB 23; Length 3664;
Best Local Similarity 79.3%; Pred. No. 83;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcaagttagcgtctgcgtttggcaagtcct 30
   ||||| |||| | ||||| |||
Db 414 tcaagaagcagcctaataatcaggcaaaagca 442

RESULT 15
AAT59975
ID AAT59975 standard; DNA; 3954 BP.
```

```

XX AC ANT59975;
XX DT 12-MAY-1997 (first entry)
XX DE
XX FE Full length tipE protein coding sequence.
XX KW Drosophila; tipE; para protein; voltage-dependent cation channel; stroke;
XX KW pesticide; insecticide; insect; parasitic infection; human; head trauma;
XX KW neuroprotection; hypoxia; therapy; ss.
XX OS
XX OS Drosophila melanogaster.
XX FH Key Location/Qualifiers
XX FT CDS 1046..2404
XX FT /*tag=a
XX FT /*product= tipE
XX PN US5593862-A.
XX PD 14-JAN-1997.
XX PF 04-OCT-1994; 94US-0317880.
XX PR 04-OCT-1994; 94US-0317880.
XX PA (UYN Y ) UNIV NEW YORK STATE RES FOUND.
XX PI Feng G, Hall LM;
XX PI WPI: 1997-099467/09.
XX DR P-PSDB; AAW13843.
XX PT Nucleic acid encoding Drosophila melanogaster tipE protein - for
XX PT prodn. of recombinant voltage-dependent cation channel
XX PS Disclosure; Column 29-32; 33pp; English.
XX CC This sequence represents the full length coding sequence for the
XX CC Drosophila tipE protein. Mutations in the tipE protein, result in a
XX CC ethyl methane sulphonate-induced recessive mutation phenotype.
XX CC Homologous flies for the mutation paralyse rapidly at 38 degrees, and
XX CC recover immediately when returned to 23 degrees. Coexpression of this
XX CC translation products that form a functional voltage-dependent cation
XX CC channel. The cation channel can be used to screen for pesticides active
XX CC against insects such as Drosophila melanogaster and pest insects. The
XX CC cation channel can also be used to screen for drugs for use in the
XX CC treatment and prevention of parasitic infections in humans and animals,
XX CC and to screen drugs for their neuroprotective effect against hypoxia,
XX CC stroke, and head trauma.
XX SQ Sequence 3954 BP; 1293 A; 960 C; 867 G; 834 T; 0 other;

Query Match 58.8%; Score 19.4; DB 18; Length 3954;
Best Local Similarity 79.3%; Pred. No. 84;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0.

Qy 2 tcaagtagcatctgtatcagcgcaagtca 30
||||| ||||| ||||| ||||| ||
Db 459 tcaagtagcagctaaatcagcgcaaaagca 487

Search completed: August 24, 2002, 22:11:54
Job time: 11964 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:55:40 ; Search time 204.64 Seconds
(without alignments)
39.611 Million cell updates/sec

Title: US-09-986-381-3
Perfect score: 33
Sequence: 1 gtcaagtagcatctgtatcaggcaagtcag 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCFUS-COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	58.8	1045	1 US-08-317-880-3	Sequence 3, Appli
2	19.4	58.8	1045	2 US-08-782-396-3	Sequence 3, Appli
3	19.4	58.8	3954	1 US-08-317-880-1	Sequence 1, Appli
4	19.4	58.8	3954	2 US-08-782-396-1	Sequence 1, Appli
5	18.6	56.4	5093	1 US-08-468-036-13	Sequence 23, Appl
6	18.6	56.4	5093	2 US-08-376-843-23	Sequence 23, Appl
7	18.2	55.2	928	3 US-08-743-975-1	Sequence 1, Appli
8	18.2	55.2	1379	2 US-08-437-607A-1	Sequence 1, Appli
9	18.2	55.2	1380	2 US-08-437-607A-1	Sequence 4, Appli
10	18.2	55.2	4621	4 US-09-592-891A-13	Sequence 5, Appli
11	17.8	53.9	2238	4 US-08-939-366-5	Sequence 7, Appli
12	17.8	53.9	2728	1 US-07-879-617A-7	Sequence 7, Appli
13	17.8	53.9	2728	1 US-08-753-985-7	Sequence 7, Appli
14	17.6	53.3	382	4 US-09-020-956-139	Sequence 139, App
15	17.6	53.3	382	4 US-09-030-607-139	Sequence 139, App
16	17.6	53.3	382	4 US-09-439-313-139	Sequence 139, App
17	17.6	53.3	1140	4 US-09-462-270-1	Sequence 1, Appli
18	17.6	53.3	1421	3 US-09-188-930-70	Sequence 70, Appl
19	17.6	53.3	1421	3 US-09-188-930-70	Sequence 254, App
20	17.6	53.3	3396	4 US-08-974-549A-640	Sequence 640, App
21	17.6	53.3	4015	4 US-08-810-009-4	Sequence 4, Appli
22	17.6	53.3	7430	4 US-08-976-259-64	Sequence 64, Appl
23	17.4	52.7	940	2 US-08-471-717-1	Sequence 1, Appli
24	17.4	52.7	6578	1 US-08-514-975B-1	Sequence 1, Appli
25	17.4	52.7	6578	5 PCT-US95-12507-1	Sequence 1, Appli
26	17.4	52.7	15222	2 US-08-801-898A-23	Sequence 23, Appl
27	17.4	52.7	15222	4 US-08-962-690-12	Sequence 12, Appl

c 28	17.4	52.7	15223	2 US-08-892-403A-1	Sequence 1, Appli
c 29	17.4	52.7	15223	4 US-08-720-132-1	Sequence 1, Appli
c 30	17.2	52.1	218	4 US-09-095-758-3	Sequence 3, Appli
c 31	17.2	52.1	218	4 US-09-422-968-3	Sequence 3, Appli
c 32	17.2	52.1	3597	4 US-09-095-758-4	Sequence 4, Appli
c 33	17.2	52.1	3597	4 US-09-422-968-4	Sequence 4, Appli
c 34	17.2	52.1	7070	1 US-08-619-554-3	Sequence 3, Appli
c 35	17.2	52.1	24417	2 US-08-846-762-1	Sequence 1, Appli
c 36	17	51.5	738	4 US-09-227-357-131	Sequence 131, App
c 37	17	51.5	2061	1 US-08-319-621A-11	Sequence 11, Appl
c 38	17	51.5	2061	1 US-08-319-621A-13	Sequence 13, Appl
c 39	17	51.5	2706	2 US-08-630-822A-61	Sequence 61, Appl
c 40	17	51.5	2706	2 US-09-005-069-61	Sequence 61, Appl
c 41	16.8	50.9	2259	1 US-07-828-700-7	Sequence 7, Appli
c 42	16.8	50.9	2313	1 US-08-232-538-5	Sequence 5, Appli
c 43	16.8	50.9	2352	1 US-08-232-538-17	Sequence 17, Appl
c 44	16.8	50.9	2352	2 US-08-786-164-17	Sequence 17, Appl
c 45	16.8	50.9	2523	4 US-09-051-363-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-317-880-3
; Sequence 3, Application US/08317880
; Patent No. 5593862
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; TITLE OF INVENTION: Proteins for Cation Channel Expression
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,880
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rogalsky, Peter
; REGISTRATION NUMBER: 38,601
; REFERENCE/DOCKET NUMBER: 19226/610 (R-5261)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1634
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-317-880-3

Query Match 58.8%; Score 19.4; DB 1; Length 1045;
Best Local Similarity 79.3%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 tcaagtagcatctgtatcaggcaagtcag 30
||||| ||||| || ||||| ||||| ||
DB 459 TCAAGAAGCAGCTAAATCAGGCAAAAGCA 487

```
RESULT 2
US-08-782-396-3
; Sequence 3, Application US/08782396
; Patent No. 5871940
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Feng, Guoping
; TITLE OF INVENTION: PROTEINS FOR CATION CHANNEL EXPRESSION
; TITLE OF INVENTION: AND FUNCTION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,396
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,880
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rogalskyj, Peter
; REGISTRATION NUMBER: 38,601
; REFERENCE/DOCKET NUMBER: 19226/611 (R-5261B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1634
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-782-396-3

Query Match 58.8%; Score 19.4; DB 2; Length 1045;
Best Local Similarity 79.3%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcaagtagcatctgtatcaggcaagtca 30
||||| ||||| || ||||| ||||| ||
Db 459 TCAAGAAGCAGCTAATCAGGCAAAAGCA 487

RESULT 3
US-08-317-880-1
; Sequence 1, Application US/08317880
; Patent No. 5593862
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Nixon, Hargrave, Devans & Doyle
; TITLE OF INVENTION: Proteins for Cation Channel Expression
; TITLE OF INVENTION: and Function
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,880
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rogalskyj, Peter
; REGISTRATION NUMBER: 38,601
; REFERENCE/DOCKET NUMBER: 19226/610 (R-5261)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1634
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3954 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-317-880-1

Query Match 58.8%; Score 19.4; DB 1; Length 3954;
Best Local Similarity 79.3%; Pred. No. 8.2;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcaagtagcatctgtatcaggcaagtca 30
||||| ||||| || ||||| ||||| ||
Db 459 TCAAGAAGCAGCTAATCAGGCAAAAGCA 487

RESULT 4
US-08-782-396-1
; Sequence 1, Application US/08782396
; Patent No. 5871940
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Feng, Guoping
; TITLE OF INVENTION: PROTEINS FOR CATION CHANNEL EXPRESSION
; TITLE OF INVENTION: AND FUNCTION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,396
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,880
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rogalskyj, Peter
; REGISTRATION NUMBER: 38,601
; REFERENCE/DOCKET NUMBER: 19226/611 (R-5261B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1634
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3954 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-782-396-1

Query Match      58.8%; Score 19.4; DB 2; Length 3954;
Best Local Similarity 79.3%; Pred. No. 8.2;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcaagtagcatctgtatcaggcaagtca 30
   ||||| ||||| || ||||| ||||| ||
Db 459 TCAAGAAGCAGCTAAATCAGGCAAAAGCA 487

RESULT 5
US-08-468-036-23/C
; Sequence 23, Application US/08468036
; Patent No. 5728806
; GENERAL INFORMATION:
; APPLICANT: DeMaggio, Anthony J.
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins that
; INTERACT WITH CASEIN KINASE I
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,036
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5728806and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-468-036-23

Query Match      56.4%; Score 18.6; DB 1; Length 5093;
Best Local Similarity 72.7%; Pred. No. 20;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaagtcatag 33
   ||||| ||||| ||||| ||||| |||||
Db 3812 GTCCAGTACCCCTGCTATTGTAAGGCTTTG 3780

RESULT 6
US-08-468-036-23

Query Match      56.4%; Score 18.6; DB 1; Length 5093;
Best Local Similarity 72.7%; Pred. No. 20;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaagtcatag 33
   ||||| ||||| ||||| ||||| |||||
Db 3812 GTCCAGTACCCCTGCTATTGTAAGGCTTTG 3780

RESULT 7
US-08-743-975-1
; Sequence 1, Application US/08743975
; Patent No. 6057434
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Mammary Transforming Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHL, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
```

```
;
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743.975
; FILING DATE: 01 NOVEMBER 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,187
; FILING DATE: 02 NOVEMBER 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-507 (PF212)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
US-08-743-975-1

Query Match 55.2%; Score 18.2; DB 3; Length 928;
Best Local Similarity 74.2%; Pred. No. 21;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtcac 31
||| ||||| ||| ||| ||||| ||| |||
Db 765 GTCTAGTAGTTCTCTGTGAAGCAAAATAAT 795

RESULT 8
US-08-437-607A-1
; Sequence 1, Application US/08437607A
; Patent No. 5955579
; GENERAL INFORMATION:
; APPLICANT: Leonard, James N. Montminy, Marc R.
; TITLE OF INVENTION: ISLET-SPECIFIC HOMEOPROTEIN AND TRANSCRIPTIONAL
; TITLE OF INVENTION: REGULATOR OF INSULIN GENE EXPRESSION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437.607A
; FILING DATE: MAY 9, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1110-1-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: HoxB13 cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 65..928
US-08-437-607A-4

Query Match 55.2%; Score 18.2; DB 2; Length 1380;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: rat CIX-1 (Hoxb13) cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..927
US-08-437-607A-1

Query Match 55.2%; Score 18.2; DB 2; Length 1379;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtcac 31
||| ||||| ||| ||| ||||| ||| |||
Db 1128 GTTCAGAGAACCTGTATCAGTCATAATCAT 1158

RESULT 9
US-08-437-607A-4
; Sequence 4, Application US/08437607A
; Patent No. 5955579
; GENERAL INFORMATION:
; APPLICANT: Leonard, James N. Montminy, Marc R.
; TITLE OF INVENTION: ISLET-SPECIFIC HOMEOPROTEIN AND TRANSCRIPTIONAL
; TITLE OF INVENTION: REGULATOR OF INSULIN GENE EXPRESSION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437.607A
; FILING DATE: MAY 9, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1110-1-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: HoxB13 cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 65..928
US-08-437-607A-4

Query Match 55.2%; Score 18.2; DB 2; Length 1380;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

; TYPE: nucleic acid
```

RESULT 11
US-08-939-366-5
Sequence 5, Application US/08939366
Patent No. 6355415
GENERAL INFORMATION:
APPLICANT: Wagner, Thomas E.
APPLICANT: Xie, Yuesheng
TITLE OF INVENTION: Compositions and Methods for the Use of
TITLE OF INVENTION: Ribozymes to Determine Gene Function
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
Zip: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/939,366
FILING DATE: 29-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OHU-02749
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410

```

; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/07/879,617A
;
; FILING DATE: 19920501
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Pabst, Patrea L.
;
; REGISTRATION NUMBER: 31,284
;
; REFERENCE/DOCKET NUMBER: EMU109
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 404-815-6508
;
; TELEFAX: 404-815-6555
;
; INFORMATION FOR SEQ ID NO: 7:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 2728 base pairs
;
; TYPE: NUCLEIC ACID
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; ORIGINAL SOURCE:
;
; ORGANISM: Rattus
;
; TISSUE TYPE: Brain
;
; IMMEDIATE SOURCE:
;
; LIBRARY: rat forebrain cDNA library
;
; CLONE: rTB2-2-20
;
; US-07-879-617A-7

```

```

Query Match      53.9%; Score 17.8; DB 1; Length 2728;
Best Local Similarity 75.9%; Pred. NO. 40;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtc 29
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GTCCAGTTGCCCTGCTGCTGCTGCAAAAGTC 184

RESULT 13
US-08-753-985-7/c
; Sequence 7, Application US/08753985
; Patent No. 5759788
; GENERAL INFORMATION:
; APPLICANT: Fremeanu Jr., Robert T.
; APPLICANT: Caron, Marc G.
; APPLICANT: Blakely, Randy D.
; TITLE OF INVENTION: A High Affinity L-Proline Transporter
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,985
; FILING DATE: 03-DEC-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879617
; FILING DATE: 01-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rattus
; TISSUE TYPE: Brain
; IMMEDIATE SOURCE:
; LIBRARY: rat forebrain cDNA library
; CLONE: rTB2-2-20
US-08-753-985-7

Query Match      53.9%; Score 17.8; DB 1; Length 2728;
Best Local Similarity 75.9%; Pred. NO. 40;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtc 29
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GTCCAGTTGCCCTGCTGCTGCTGCAAAAGTC 184

Query Match      53.3%; Score 17.6; DB 4; Length 382;
Best Local Similarity 71.9%; Pred. NO. 32;
Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtcata 32
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 GTCCGGTAGCACCTGAGTAAGCAAAATGCAGA 176

RESULT 14
US-09-020-956-139/c
; Sequence 139, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 139:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-020-956-139

Query Match      53.3%; Score 17.6; DB 4; Length 382;
Best Local Similarity 71.9%; Pred. NO. 32;
Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtcata 32
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 GTCCGGTAGCACCTGAGTAAGCAAAATGCAGA 176

RESULT 15
US-09-030-607-139/c
; Sequence 139, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
US-09-030-607-139/c
```


SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030.607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 139:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-139

Query Match 53.3%; Score 17.6; DB 4; Length 382;
Best Local Similarity 71.9%; Pred. No. 32;
Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gtcaagtagcatctgtatcaggcaaaagtcata 32
||| ||||| ||| | ||||| |||
Db 207 GTCCGGTAGCACCTGAGTAGGCAAAATGCAGA 176

Search completed: August 24, 2002, 21:55:43
Job time: 13138 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2002, 21:08:07 ; Search time 7654.23 seconds
(without alignments)
58.190 Million cell updates/sec

Title: US-09-986-381-3

Perfect score: 33
Sequence: 1 gtaagtagcatctgtatcaggaagtcataag 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estro:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22.4	67.9	345	10	BE765117
C 2	21.8	66.1	216	12	AZ797581
C 3	21.4	64.8	486	12	AQ792494
C 4	21.4	64.8	541	10	BI467483
C 5	21.4	64.8	759	12	AZ751369
C 6	21.4	64.8	933	10	BF966971
C 7	20.8	63.0	483	9	AI440010
C 8	20.8	63.0	721	12	BH347092
C 9	20.8	63.0	787	12	AQ855061
C 10	20.6	62.4	443	12	AQ973627
C 11	20.6	62.4	478	12	AZ172308
C 12	20.6	62.4	506	12	BH281027
C 13	20.4	61.8	53	10	R37946
C 14	20.4	61.8	153	9	AW579897
C 15	20.4	61.8	252	10	BF882238
C 16	20.4	61.8	266	9	AW270798
C 17	20.4	61.8	273	9	AA385837

C 18	20.4	61.8	277	10	F04554
C 19	20.4	61.8	278	9	BE082087
C 20	20.4	61.8	296	9	AW062718
C 21	20.4	61.8	314	9	BE082139
C 22	20.4	61.8	340	10	F00612
C 23	20.4	61.8	352	10	BF930784
C 24	20.4	61.8	355	10	BF801238
C 25	20.4	61.8	355	10	BF326991
C 26	20.4	61.8	359	10	H88591
C 27	20.4	61.8	367	9	AI492621
C 28	20.4	61.8	368	10	BE769126
C 29	20.4	61.8	370	10	BE709117
C 30	20.4	61.8	374	10	BE769180
C 31	20.4	61.8	377	10	BE814128
C 32	20.4	61.8	382	9	AW889584
C 33	20.4	61.8	406	10	D53080
C 34	20.4	61.8	412	9	AW803629
C 35	20.4	61.8	412	9	BI043756
C 36	20.4	61.8	424	9	AV733123
C 37	20.4	61.8	424	10	BF349170
C 38	20.4	61.8	445	10	BF056800
C 39	20.4	61.8	452	9	AW197646
C 40	20.4	61.8	456	9	AV699107
C 41	20.4	61.8	456	10	F37528
C 42	20.4	61.8	458	9	AI766287
C 43	20.4	61.8	470	10	BI438920
C 44	20.4	61.8	470	10	BE775114
C 45	20.4	61.8	472	9	AI074492

ALIGNMENTS

RESULT 1

BE765117/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BE765117 346 bp mRNA linear EST 19-SEP-2000
CMI-NT0088-300500-232-g03 NT0088 Homo sapiens cDNA, mRNA sequence.
BE765117
BE765117.1 GI:10194950
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl1&t2=CM1-NT0088-300
500-232-g03&t3=2000-05-30&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 69
High quality sequence stop: 346.
Location/Qualifiers
1. .346
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="NT0088"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
101 a 74 c 60 g 111 t
BASE COUNT
ORIGIN

Query Match 67.9%; Score 22.4; DB 10; Length 346;
Best Local Similarity 81.2%; Pred. No. 65;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcaagtacatctgtatcaggcaaatgcatag 33
||||| ||||||| ||||||| |||||||
Db 172 TCAAGTAACATCTGTGTCAGGCCATCATAG 141

RESULT 2
LOCUS Az797581 216 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0053D23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0053D23 R, DNA sequence.
ACCESSION Az797581
VERSION Az797581.1 GI:12946808
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 216)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0053 row: D column: 23
Seq primer: CACACAGGAACACTATGACC
Class: plasmid ends
High quality sequence stop: 216.
Location/Qualifiers
1..216
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0053D23"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
50 a 38 c 58 g 70 t
BASE COUNT
ORIGIN

Query Match 66.1%; Score 21.8; DB 12; Length 216;
Best Local Similarity 78.8%; Pred. No. 96;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 gtcaagtagcatctgtatcaggcaaatgcatag 33
||||| ||||||| ||||||| |||||||
Db 87 GTCCAGTAGCATCTGTATCGTATAGAAATGG 119

RESULT 3
LOCUS A0792494 486 bp DNA linear GSS 03-AUG-1999
DEFINITION HS_5451_A1_E06_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1027 Col=11 Row=I, DNA sequence.
ACCESSION A0792494
VERSION A0792494.1 GI:5700118
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 486)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1027 row: I column: 11
Seq primer: T7
Class: BAC ends
High quality sequence stop: 486.
Location/Qualifiers
1..486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1027 Col=11 Row=I"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methyase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
FEATURES
source

```

RESULT 5

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10040 row: a column: 01
 High quality sequence start: 112
 High quality sequence stop: 903.

FEATURES

Location/Qualifiers
 1. .933

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4375176"
 /clone_lib="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTITVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

248 a 267 c 168 g 250 t

BASE COUNT

ORIGIN

Query Match 64.8%; Score 21.4; DB 10; Length 933;
 Best Local Similarity 80.6%; Pred. No. 2.3e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 caagtagcatctgtatcaggcaagtcataag 33

||||| ||||||| ||||||| |||||||

Db 200 CAAGTAACATCTGTATCAGCGCGCTACATAG 170

RESULT 7

AI440010/c

LOCUS

DEFINITION
 AI440010 483 bp mRNA linear EST 09-MAR-1999
 t163h06.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2135195 3'
 similar to gb:L12693 CELLULAR NUCLEIC ACID BINDING PROTEIN (HUMAN
); mRNA sequence.

ACCESSION
 AI440010

VERSION
 AI440010.1 GI:4308162

KEYWORDS
 EST.

SOURCE
 human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 483)

NCI_CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Life Technologies catalog #: 11547-015

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 286.

Location/Qualifiers

1. .483

/organism="Homo sapiens"

/db_xref="taxon:9606"

FEATURES

source

/clone="IMAGE:2135195"
 /clone_lib="NCI_CGAP_Lym12"
 /tissue_type="lymphoma, follicular mixed small and large
 cell"

/lab_host="DH10B"

/note="Organ: lymph node; Vector: pCMV-SPORT6; Site 1:

SalI; Site 2: NotI; Cloned unidirectionally. Primer:

Oligo dT. Average insert size 1.25 kb. Life Technologies

catalog #: 11547-015"

134 a 98 c 70 g 180 t 1 others

BASE COUNT

ORIGIN

Query Match 63.0%; Score 20.8; DB 9; Length 483;

Best Local Similarity 88.0%; Pred. No. 3.1e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tcaagtagcatctgtatcaggcaaaa 26

||||| ||||||| ||||||| |||||||

Db 448 TCAAGTAACATCTGTATCAGCGNTA 424

RESULT 8

BH347092/c

LOCUS

DEFINITION
 BH347092 721 bp DNA linear GSS 03-DEC-2001
 CH230-54E9.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

ACCESSION
 BH347092

VERSION
 BH347092.1 GI:17277826

KEYWORDS
 GSS.

SOURCE
 Norway rat.

ORGANISM
 Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 721)

Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn

,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de

Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other GSSs: CH230-54E9.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/orering_information.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 54 row: E column: 9

Seq primer: T7

Class: BAC ends.

FEATURES

source

1. .721

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-54E9"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

236 a 141 c 113 g 231 t

BASE COUNT

ORIGIN

RESULT 10

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 478)
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.

A sea urchin genome project: Sequence scan, virtual map, and additional resources

JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566

Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 121 row: F column: 4

Seq primer: T7

Class: BAC ends

High quality sequence stop: 478.

Location/Qualifiers

1..478

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone_xref="plate:121 Col=4 Row=F"

/clone_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"

/note="organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli

DH10B"

BASE COUNT 125 a 97 c 84 g 163 t 9 others

ORIGIN

Query Match 62.4%; Score 20.6; DB 12; Length 478;

Best Local Similarity 85.2%; Pred. No. 3.8e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 tagcatctgtatcaggcaaaagtcataag 33

II IIIIIIIII IIIIIIIIIII

Db 295 TAACATCTGTAACATGCAGAAAGTCATGG 269

RESULT 12

BH281027/c

LOCUS

DEFINITION

CH230-38M12.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone

BH281027

CH230-38M12, DNA sequence.

ACCESSION

BH281027.1 GI:17193429

KEYWORDS

GSS.

SOURCE

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 506)

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn

, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de

Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 38 row: M column: 12

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..506

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-38M12"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: PTARBAC2.1; Site1: EcoRI; Site2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

BASE COUNT 144 a 119 c 117 g 126 t

ORIGIN

Query Match 62.4%; Score 20.6; DB 12; Length 506;

Best Local Similarity 85.2%; Pred. No. 3.8e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 gtacatctgtatcaggcaaaagtcata 32

IIIIIIIIII IIIIIIIIIII

Db 208 GTAGCATCTGTGTCATGTAAAGGCATA 182

RESULT 13

R37946/c

LOCUS

DEFINITION

YC59C07.sl Soares infant brain LNIB Homo sapiens cDNA clone

IMAGE:24245 3' similar to gb:U12693 CELLULAR NUCLEIC ACID BINDING

PROTEIN (HUMAN); mRNA sequence.

ACCESSION

R37946

VERSION

R37946.1 GI:795402

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 53)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

, R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1267

High quality sequence starts: 1 High quality sequence stops: 1

Source: IMAGE Consortium, LNIB. This clone is available royalty-free

through LNIB; contact the IMAGE Consortium (info@image.llnl.gov)

for further information. Trace considered overall poor quality

Insert Length: 1267 Std Error: 0.00

Seq primer: -21ml3

High quality sequence stop: 1.

Location/Qualifiers

1..53

/organism="Homo sapiens"

/db_xref="GDB:396592"

/db_xref="taxon:9606"

/clone="IMAGE:24245"

/clone_lib="Soares infant brain LNIB"

FEATURES

source


```

/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lfamid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AACGTGGAAGAAATCGCGCGCAGGAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      16 a  9 c  11 g  15 t  2 others
ORIGIN

```

```

Query Match      61.8%; Score 20.4; DB 10; Length 53;
Best Local Similarity 95.5%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 2 tcaagtagcatctgtatcaggc 23
    ||||||| ||||||| |||||||
Db 23 TCAAGTAACATCTGTATCAGGC 2

```

```

RESULT 14
LOCUS      AW579897      153 bp      mRNA      linear      EST 16-MAR-2000
DEFINITION OVO-HT0368-040100-082-h01 HT0368 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW579897
VERSION     AW579897.1 GI:7255050
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 153)
            HCGP http://www.ludwig.org.br/ORESTES.
AUTHORS    The FAPESP/LICR Human Cancer Genome Project
JOURNAL    Unpublished (1999)
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=QV0&st2=QV0-HT0368-040100-082-h01&st3=2000-01-04&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 153.

```

```

FEATURES
Source
1..153
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="HT0368"
    /dev_stage="Adult"
    /note="Organ: head neck; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT      48 a  27 c  26 g  51 t  1 others
ORIGIN

```

```

Query Match      61.8%; Score 20.4; DB 9; Length 153;

```

```

Best Local Similarity 95.5%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcaagtagcatctgtatcaggc 23
    ||||||| ||||||| |||||||
Db 70 TCAAGTAACATCTGTATCAGGC 49

```

```

RESULT 15
LOCUS      BF882238/c      252 bp      mRNA      linear      EST 17-JAN-2001
DEFINITION QV1-ET0183-051200-547-b04 ET0183 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF882238
VERSION     BF882238.1 GI:12272364
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 252)
AUTHORS    Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
            Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
            Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
            Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
            , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
            Simpson, A.J.

```

```

TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=QV1&st2=QV1-ET0183-051200-547-b04&st3=2000-12-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 252.

```

```

FEATURES
Source
1..252
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="ET0183"
    /dev_stage="Adult"
    /note="Organ: lung tumor; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT      75 a  60 c  41 g  76 t
ORIGIN

```

```

Query Match      61.8%; Score 20.4; DB 10; Length 252;
Best Local Similarity 95.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 2 tcaagtagcatctgtatcaggc 23
    ||||||| ||||||| |||||||
Db 64 TCAAGTAACATCTGTATCAGGC 43

```

```

Search completed: August 24, 2002, 21:08:14
Job time: 17334 sec

```

